

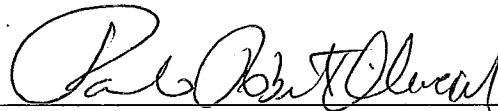
MÉTODOS GEODÉSICOS NA PROGRAMAÇÃO MATEMÁTICA

por

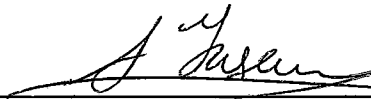
JOÃO XAVIER DA CRUZ NETO

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO EM ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

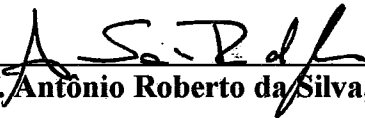
Aprovado por:



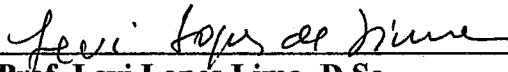
**Prof. Paulo Roberto Oliveira, Dr. Ing.
(Presidente)**



Prof. Alfredo Noel Iusem, Ph.D.



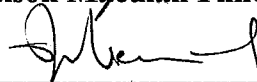
Prof. Antônio Roberto da Silva, D.Sc.



Prof. Levi Lopes Lima, D.Sc.



Prof. Nelson Maculan Filho, D.Sc.



Profa. Suzana Scheimberg de Makler, D.Sc.

SETEMBRO / 95

Neto, João Xavier da Cruz

Métodos Geodésicos na Programação

Matemática [Rio de Janeiro] 1995.

IIX 0, 102 p. 29,7 cm (COPPE/UFRJ, D.Sc.,

Engenharia de sistemas e Computação, 1995)

Tese-Universidade Federal do Rio

de Janeiro, COPPE.

1. Otimização

I. COPPE/UFRJ II. Título (série)

Dedico à:
memória de meu pai,
Paulo Leite
minha mãe, Maria
Fernandes
minha esposa, Ana Zélia
minha filha, Mariana
meu filho, Mateus.

AGRADECIMENTOS

Ao Prof. Paulo Roberto pela sugestão do tema e pelas palavras motivadoras durante toda minha permanência na UFRJ.

Aos amigos Orizon, Quirino, Marcos Augusto, Benar e Alfredo pelo convívio e inúmeras discussões sobre os mais variados temas.

Ao Departamento de Matemática da Universidade Federal do Piauí; em especial aos amigos Gilvan e Barnabé.

Ao casal de amigos, Cícero-Ivanilde, pela convivência durante os últimos cinco meses no Rio.

À minha família pelo apoio e estímulo durante toda travessia do doutorado.

Finalmente, à Wilson Góes e José Pio pela grande ajuda na digitação eletrônica desta tese.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para obtenção do grau de Doutor em Ciências (M.Sc.).

MÉTODOS GEODÉSICOS NA PROGRAMAÇÃO MATEMÁTICA

João Xavier da Cruz Neto

SETEMBRO, 1995

Orientador: Prof. Paulo Roberto Oliveira

Programa: Engenharia de Sistemas e Computação

Área: Otimização

O Método de Descida Geodésica foi proposto por Luenberger no ano de 1972 com o objetivo de obter uma taxa assintótica de convergência para o Método de Gradiente Projetado, que resolve o problema de minimizar $f: \mathbf{R}^n \rightarrow \mathbf{R}$ sobre a restrição $h(x) = 0 (h: \mathbf{R}^n \rightarrow \mathbf{R}^m)$.

Nesta tese usamos as idéias de Luenberger para obter algoritmos conceituais na resolução do problema $\min f(x)$, $x \in M$ e M variedade riemanniana completa.

Desenvolvemos o conceito de hessiano de uma função restrita a uma subvariedade e derivamos condições de otimalidade que permitem obter de maneira simples os resultados clássicos.

Desenvolvemos elementos de Análise Convexa em Variedades Completas que permitem obter resultados de convergência global para minimização de funções convexas em variedades de curvatura não-negativa.

Fazemos uma relação entre métricas, definida no conjunto de restrições, e direções de descida usada por diversos algoritmos. Isto permite mostrar, por exemplo, que o campo de direções usado no algoritmo Projetivo de Karmarkar é um campo gradiente.

Summary of the PHD thesis presented to COPPE/UFRJ as a part of the necessary requisits needed to obtain the degree of Doctor in Sciences.

GEODESIC METHODS IN MATHEMATICAL PROGRAMMING

João Xavier da Cruz Neto

SEPTEMBER, 1995

Advisor: Prof. Paulo Roberto Oliveira

Program: System Engineering and Computation

Field: Optimization

The Geodesic Descent Method was proposed by Luenberger in 1972 with the aim of obtaining an assymptotic convergence rate for the projected gradient method, which solves the problem of minimizing $f: \mathbf{R}^n \rightarrow \mathbf{R}$ under the constraint $h(t) = 0$ ($h: \mathbf{R}^n \rightarrow \mathbf{R}$).

In this thesis we used Luenberger's idea in order to obtain conceptual algorithms for the resolution of the problem $\min f(x)$, $x \in M$ where M is a complete Riemannian manifold.

We develop the concept of the hessian of a function restricted for a submanifold and we derived optimality conditions that allows one to obtain in a simple way the classical results.

We developed the elements of convex analysis in complete manifolds, which allows the obtention of global convergence results for minimizing convex function on manifolds of non-negative curvature.

We established a relation between metrics in the feasible set and descent directions employed by many algorithms. This allows one to show that, for instance, the direction field employed in Karmarkar Projective Algorithm is a gradient field.

SUMÁRIO

Introdução	1
Capítulo 1 – Elementos de Geometria Riemanniana	6
1. Variedades Diferenciáveis em \mathbf{R}^n	6
2. Espaço Tangente a uma Variedade	7
3. Métrica Riemanniana	9
4. Grupo de Lie e Métrica Invariante	12
5. Conexões Afins e Derivada Covariante	19
6. Campos Paralelos e Geodésicas	23
7. Conexão Riemanniana	24
8. Curvaturas	30
9. Subvariedade	35
10. Gradiente e Hessiana	39
11. Variedade Completa e Teorema de Hopf/Rinow	50
12. Teorema de Topogonov e a Lei dos Cossenos	54
Capítulo 2 – Fundamentos: Condições de Otimalidade e elementos de Análise Convexa	58
1. Condições de Otimalidade	58
2. Convexidade em Variedade Riemanniana	62
Capítulo 3 – Métodos Geodésicos	68
1. Uma Visão de Métodos Primais Através de Métricas Riemannianas	68
2. Métodos de Descida Geodésica	76
3. Exemplo	81
4. Convergência Global: Problemas Convexos em Variedades de Curvatura Não-Negativa	85

Conclusão	90
Bibliografia	92

MÉTODOS GEODÉSICOS NA PROGRAMAÇÃO MATEMÁTICA

INTRODUÇÃO

As relações da Geometria Diferencial (GD) e Cálculo Variacional são explícitas, por exemplo, na obtenção das superfícies de revolução de área mínima, e no estudo de geodésicas.

A discretização de tais problemas pode levar a modelos de Programação Não-Linear (PNL), filha em dimensão finita do Cálculo Variacional.

Há entretanto uma outra via que relaciona a PNL, assim como a Programação Linear (PL) à Geometria, que é tomada como instrumento de análise dos métodos existentes naquelas áreas de Programação Matemática.

A nosso conhecimento, o primeiro trabalho com esta perspectiva é o de Luenberger [72] que obtém, através do conceito de geodésica, a taxa de convergência do método de gradiente projetado aplicado ao problema $\min f(x)$, s.a. $h(x) = 0$, onde $f: \mathbf{R}^n \rightarrow \mathbf{R}$, $h: \mathbf{R}^n \rightarrow \mathbf{R}^m$, $m < n$.

Luenberger propõe inicialmente um método que denomina de “descida geodésica” (veja Figura 1):

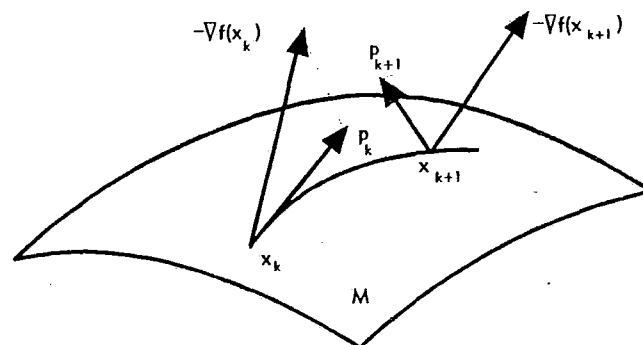


Figura 1

1. dado x_k viável ($h(x_k) = 0$), calcule a projeção p_k de $-\nabla f(x_k)$ no plano tangente à superfície das restrições $M := \{x: h(x) = 0\}$;
2. determine a geodésica (única) $x(t)$, $t \geq 0$, de M , que verifique $x(0) = x_k$ e $\dot{x}(0) = p_k$;
3. minimize $f(x(t))$, $t \geq 0$, obtendo t_k , e $x_{k+1} = x(t_k)$.

A idéia subjacente à utilização da descida geodésica como uma aproximação do gradiente projetado é o fato das geodésicas se comportarem localmente como os segmentos que seriam obtidos pela projeção do gradiente se M fosse plano. Além disto, à medida que o processo converge, a distância entre os pontos obtidos por cada um dos dois métodos tende para zero (segundo Luenberger, mais rapidamente que os respectivos passos t_k). Assim, as taxas assintóticas de convergência de ambos os métodos é a mesma. Além da taxa assintótica de convergência, Luenberger estabelece a convergência global do método sobre hipóteses razoáveis:

Teorema de Convergência Global (Luenberger [72]):

Sejam $f: \mathbf{R}^n \rightarrow \mathbf{R}$, $h: \mathbf{R}^n \rightarrow \mathbf{R}^m$, $m < n$, funções de classe C^3 . Suponha que f é limitada inferiormente e que "0" é valor regular para a função h . Para cada $x \in M = \{x \in \mathbf{R}^n; h(x) = 0\}$ defina

$$\lambda(x) = \nabla f(x) \nabla h^T(x) [\nabla h(x) \nabla h(x)^T]^{-1}$$

e suponha que

$$L(x) = \nabla^2 f(x) + \sum_{i=1}^m \lambda_i(x) \nabla^2 h_i(x)$$

satisfaz $aI \leq L(x) \leq AI$, $a > 0$ em $T_x M$ (espaço tangente de M em x). Se $\{x_k\}$ é a seqüência de pontos gerada pelo método de descida geodésico, então a correspondente seqüência do negativo dos gradientes projetados $p_k = -\nabla f(x_k) - \lambda(x_k) \nabla h(x_k)$ converge a zero. Além disso, $x_k \rightarrow x_0 \in M$ que minimiza f sobre M .

Gabay [82] propôs esta metodologia para os métodos de Newton e Quase-Newton, obtendo teoria e taxa assintótica de convergência similares aos casos irrestritos do \mathbf{R}^n ; além de trabalhar em um contexto mais geral, $M \subset \mathbf{R}^n$ variedade diferenciável munida da métrica induzida do espaço euclidiano, estabeleceu elementos de condições de otimalidade.

Outra extensão do trabalho de Luenberger encontra-se em Rapsak [91], no qual estabelece alguns elementos de análise convexa no contexto considerado por Gabay.

Uma outra linha de pesquisa tem sido objeto de atuação de diversos autores, que têm analisado, através de instrumentos da Geometria Diferencial, as propriedades das trajetórias de algoritmos de pontos interiores da Programação Linear. Citemos os essenciais Bayer-Lagarias [89a, 89b], Lagarias [90] e Karmarkar [90]. O objetivo de todos eles é a análise das trajetórias contínuas associadas a determinadas classes de métodos. Suponha um problema genérico,

$$\min f(x), \quad x \in M \subset \mathbf{R}^n. \quad (1)$$

Considere um método destinado à resolução de (1) dado por:

$$x_{k+1} = x_k + t_k d_k, \quad k \geq 0, \quad x_0 \text{ dado.} \quad (2)$$

Em (2), d_k é uma direção em \mathbf{R}^n , em geral de descida para alguma função φ associada a (1), isto é, $\varphi(x_k + td_k) < \varphi(x_k)$, para $t > 0$ suficientemente pequeno; $t_k > 0$ é o passo dado nesta direção.

Considere o limite em (2) para $t_k \rightarrow 0$; obtém-se a equação diferencial

$$\frac{dx}{dt} = d.$$

Como d é uma função de x , ela define um campo de vetores definido em M . A representação completa do problema é

$$\frac{dx(t)}{dt} = d(x(t)), \quad x(0) = x_0. \quad (3)$$

$x(t)$, $t \geq 0$ é a trajetória contínua associada ao algoritmo discreto (2). Sua análise em $M \subset \mathbf{R}^n$ dependerá, é claro, da expressão de d , mas também das propriedades de M .

Este tipo de análise não tem valor apenas teórico. Questões como curvatura de $x(t)$ e complexidade do cálculo de $\{x_k\}$ são relacionadas, como demonstrou Karmarkar em [90]. Os trabalhos de Meggido-Shub [89] e Adler-Monteiro [90] derivam propriedades para a seqüência

gerada por (2) a partir de propriedades do sistema dinâmico (3) associadas ao algoritmo afim-escala de Dikin [67] e ao algoritmo projetivo de Karmarkar [84].

A tese é dividida em três capítulos.

O Capítulo 1 tem como objetivo principal desenvolver um texto com elementos de Geometria Riemanniana necessários à leitura dos vários artigos citados, veja Oliveira-Xavier [94], bem como servir de subsídios aos capítulos posteriores. Ao mesmo tempo em que expomos os conceitos de Geometria, veja do Carmo [88], desenvolvemos exemplos ilustrativos de interesse da Programação Matemática sempre procurando adaptá-los aos objetivos dos capítulos seguintes. Como exemplo, citemos as expressões explícitas para geodésicas, gradientes e hessianos no ortante positivo e no simplex unitários, dois modelos de bastante interesse para a Programação Matemática. A dedução de expressões para o hessiano de funções restritas a uma subvariedade, por exemplo $M = h^{-1}(0) \subset \mathbf{R}^n$, o uso do Teorema de Hopf-Rinow nas provas de convergência de algoritmos de descida geodésica e uma “lei dos cossenos” (Teorema 1.14) para variedades completas de curvatura não-negativa são contribuições importantes do capítulo.

No Capítulo 2, usamos o material desenvolvido na seção 1.11, estabelecendo condições de Otimalidade num contexto amplo, que inclui o apresentado por Gabay [82] e, em particular, deduzimos o Teorema dos Multiplicadores de Lagrange (Condição Necessária de 1ª Ordem) e a condição de Semi-Definição Positiva do Hessiano do Lagrangeano, restrito ao espaço tangente das restrições (Condição Necessária de 2ª Ordem). Apresentamos ainda elementos de Análise Convexa e estabelecemos Condições Necessárias e Suficientes para Convexidade de Funções; neste contexto, o Teorema de Convergência Global de Luenberger pode ser visto como estabelecendo a convergência do método gradiente para funções fortemente convexas definidas em $M = h^{-1}(0)$. Outro resultado interessante é o Corolário 2.6, que para

$f: M \rightarrow \mathbf{R}$ convexa e M compacta, demonstra que f é constante.

Observe que se h é tal que $h^{-1}(0)$ é compacta, então, no Teorema de Luenberger, não existiria função f com tal hipótese.

No Capítulo 3 tratamos de dois problemas: o primeiro consiste em, dado o Sistema Dinâmico (3) associado ao algoritmo (2), determinar a existência de uma métrica no conjunto de restrições de modo que as curvas integrais de (3) correspondam às curvas integrais do campo gradiente da função objetivo (a ser minimizada). Na resolução do problema de Programação Linear com o método de Dikin, Karmarkar forneceu uma resposta positiva, e está apresentado em Bayar-Lagorias [89]. Fornecemos uma outra demonstração para o resultado acima e mostramos que o Campo Projetivo de Karmarkar é uma representação do campo de direções de Dikin no simplex unitário, portanto, campo gradiente; além disso, introduzindo a métrica definida pelo hessiano da função $h(x) = \sum_{i=1}^n x_i \log x_i$ no ortante positivo, mostramos que o sistema (3) associado ao algoritmo multiplicativo, proposto por Eggermont [90] (convergência estudada por Iusem [95]), corresponde às curvas integrais do campo gradiente da função objetivo. O segundo problema trata de estabelecer uma teoria de convergência do método gradiente com buscas exata e inexata. Com a hipótese sobre a função $f: \mathbf{R}^n \rightarrow \mathbf{R}$ de ser de classe C^1 , com gradiente Lipschitziano e existência de mínimo, Burachik [95] estabeleceu a convergência global da seqüência escolhendo a priori o tamanho do passo λ_k em um intervalo que depende da constante de Lipschitz do gradiente. A demonstração se baseia na convergência quasi-Féjer e na desigualdade

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 + \lambda_k^2 \|\text{grad } f(x)\|^2 - 2\lambda_k(f(x_k) - f(x^*)) \quad (4)$$

onde $\{x_k\}$ é a seqüência gerada pelo algoritmo e x^* é um ponto de mínimo. Kiwiel [83] obteve, em primeiro a expressão (4) em \mathbf{R}^n . Com a hipótese de curvatura não-negativa, generalizamos a desigualdade (4) (Teorema 3.13), onde a distância euclidiana é substituída pela distância intrínseca da variedade. Com isso, estabelecemos a convergência global da seqüência gerada pelo método de descida geodésica para variedades de curvatura não-negativa.

De um modo geral, não demonstramos os resultados contidos na literatura citada; portanto, o que se encontra demonstrado é de nossa autoria ou é um resultado de fácil dedução da teoria desenvolvida e, possivelmente, com uma demonstração diferente.

CAPÍTULO 1

ELEMENTOS DE GEOMETRIA RIEMANNIANA

1.1. INTRODUÇÃO

Neste capítulo apresentamos elementos de Geometria Riemanniana que serão usados nos Capítulos 2 e 3; os resultados apresentados, sem demonstrações, encontram-se em Do Carmo [88] e ilustramos com exemplos da Programação Matemática.

1.2. Variedades Diferenciais em \mathbf{R}^n

Limitâmo-nos à apresentação das idéias básicas que definem uma variedade. É um conjunto M em \mathbf{R}^n formado pela reunião de um número finito ou enumerável de domínios M_j ; cada M_j possui um sistema de coordenadas, denominadas locais, que de modo único determinam qualquer ponto de M_j por, digamos, (x_j^1, \dots, x_j^n) ; se dois domínios M_j e M_k se interceptarem, os pontos comuns terão coordenadas de M_j (resp. M_k) que podem ser expressas por uma função diferenciável das coordenadas de M_k (resp. M_j). Pode-se supor que estas funções de passagens são de classe C^∞ .

Exemplo 1.1: $M \equiv \mathbf{R}^n$

Exemplo 1.2: $M \equiv \{(x, y); y > 0\} \subset \mathbf{R}^2$

Exemplo 1.3: $M \equiv \text{int } \mathbf{R}_+^n := \{x \in \mathbf{R}^n; x_i > 0\}$

Exemplo 1.4: $M \equiv f^{-1}(c)$, onde $f: D \subset \mathbf{R}^n \rightarrow \mathbf{R}$, D é um conjunto aberto de \mathbf{R}^n , f é diferenciável e $\nabla f(x) \neq 0, \forall x \in D$.

Em \mathbf{R}^2 , M corresponde às curvas de nível de f .

Exemplo 1.5: Se $f(x) = x_1^2 + \dots + x_n^2$, $M = f^{-1}(1)$ é a “esfera” de \mathbf{R}^n centrada na origem, de raio 1. Observe que apenas um sistema de coordenadas locais é insuficiente para determinar os pontos de M .

Exemplo 1.6: Se $f(x) = a^T x$, com $\beta \in \mathbf{R}$, $a \neq 0$, $M = f^{-1}(\beta)$ é o hiperplano de \mathbf{R}^n , $a^T x = \beta$.

Exemplo 1.7: Se $f(x) = e^T x$, onde $e = (1, \dots, 1)^T$, e $f: D \rightarrow \mathbf{R}$, onde $D = \text{int } \mathbf{R}_+^n$, então $M = f^{-1}(1) = \{x : e^T x = 1, x > 0\}$.

1.3. Espaço Tangente a uma Variedade – Fibrado Tangente

Os exemplos que acabamos de apresentar são convincentes da importância da estrutura de variedade diferenciável. Na verdade, ela é ainda muito ampla para nossos objetivos. Vamos, neste número e a seguir, enriquecê-la com outros conceitos, que serão também aplicáveis aos exemplos citados.

Em uma variedade diferenciável pode-se definir curvas suaves (parametrizadas) dadas por $\alpha: I \rightarrow M$, onde I é algum intervalo de \mathbf{R} , $\alpha(t) = (\alpha_1(t), \dots, \alpha_n(t))$. Se t representar o tempo, a derivada é o vetor velocidade, tangente à curva, dado por $\dot{\alpha}(t) = (\dot{\alpha}_1(t), \dots, \dot{\alpha}_n(t))$, aplicado em $\alpha(t)$ (veja Figura 2).

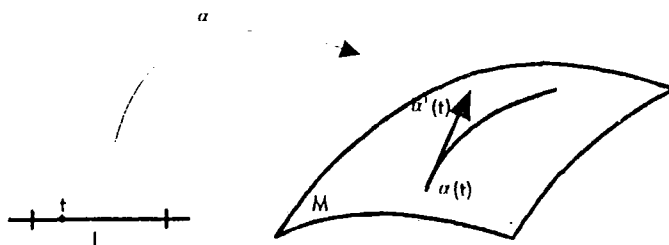


Figura 2

Tomamos como definição de espaço tangente a uma variedade M em um ponto p o conjunto de todos os vetores tangentes às curvas suaves pertencentes a M passando por p ; representamo-lo por $T_p M$.

Conforme veremos através de exemplos, a variação contínua de p em M leva a uma variação contínua, de fato diferenciável, de $T_p M$. A união destes espaços tangentes $\bigcup_{p \in S} T_p M$ é o que se denomina de fibrado tangente. Observe que neste conceito está explícito o sistema de coordenadas locais anteriormente descrito.

Os exemplos a seguir se referem aos de mesmo número de variedades diferenciais:

Exemplo 1.1: $T_p M \equiv M = \mathbf{R}^n, \forall p$.

Exemplo 1.2: $T_p M \equiv \mathbf{R}^2, \forall p$.

Exemplo 1.3: $T_p M \equiv \mathbf{R}^n, \forall p$.

Exemplo 1.4: $T_p M \equiv \{\nabla f(p)^\perp\}$ (veja Figura 3)

Exemplo 1.5: $T_p M \equiv \{(2x_1, \dots, 2x_n)^\perp\}^\perp$ (veja Figura 3) ($x = (x_1, \dots, x_n)$ representa p).

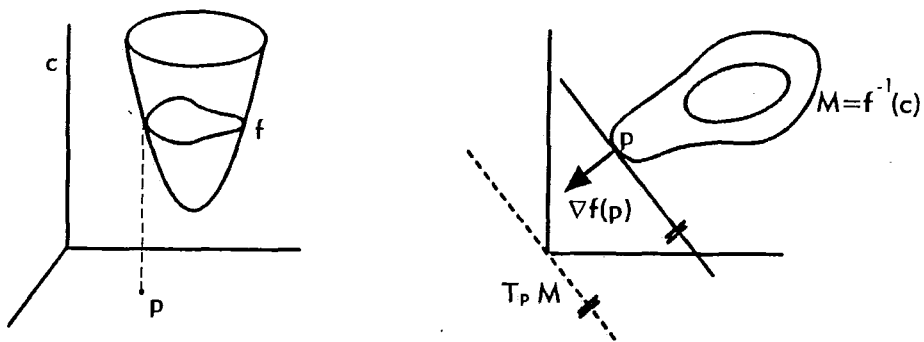


Figura 3

Exemplo 1.6: $T_p M \equiv a^\perp$ (veja Figura 4).

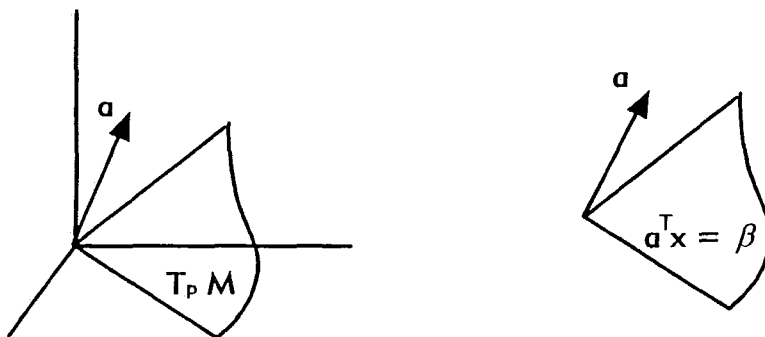


Figura 4

Exemplo 1.7: $T_p M \equiv \{e^\perp\}$ (veja Figura 5)

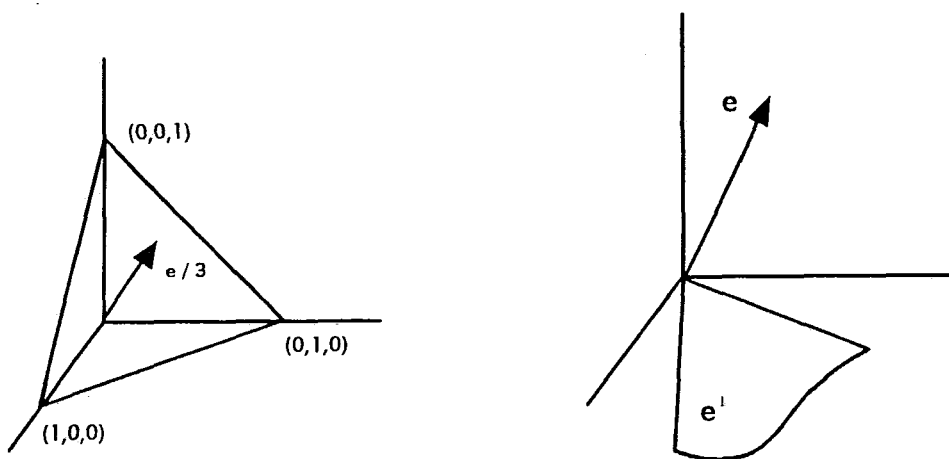


Figura 5

1.4. Métricas Riemannianas em Variedades Diferenciáveis

Em espaços lineares, como \mathbf{R}^n , ou de matrizes, ou de funções, estamos habituados ao essencial conceito de métrica, necessário para determinar distância entre pontos do espaço, comprimentos de curvas, análise de erros, etc.

Uma forma de estender este conceito a Variedades é observar sua aplicação a curvas. Uma curva em \mathbf{R}^n pode ser representada parametricamente por $x_1 = f_1(t), \dots, x_n = f_n(t)$, $t \in [a, b]$, onde as funções f_i são diferenciáveis. O vetor $v(t) = \left(\frac{df_1}{dt}, \dots, \frac{df_n}{dt} \right)$ é o vetor velocidade, tangente à curva em t . O comprimento da curva, obtido por cálculo elementar é dado por

$$\ell = \int_a^b \sqrt{\sum_i \left(\frac{df_i}{dt} \right)^2} dt = \int_a^b \sqrt{(v(t), v(t))} dt = \int_a^b |v(t)| dt,$$

onde $|\cdot|$ representa a norma euclidiana do vetor. O comprimento ℓ é portanto a integral do comprimento do vetor tangente à curva. Fica assim claro que se uma curva pertencer a uma variedade diferenciável M , seu comprimento será obtido por medida realizada em vetores pertencentes ao espaço tangente à curva em cada ponto. Esta medida é o produto escalar $(v, v)_t$ calculado no ponto t da curva.

Necessitamos portanto definir um produto escalar $(\cdot, \cdot)_p$ no espaço tangente $T_p M$, para todo $p \in M$. De um modo geral, sabemos que além do produto euclidiano tradicional, $(v, w)_p = \sum_i v_i(p)w_i(p)$, podemos definir $(v, w)_p = \sum_{i,j} g_{ij} v_i(p)w_j(p) = (Gv, w)_p$, onde G é uma matriz simétrica definida positiva. Esta definição aparece de modo natural ao realizarmos uma mudança de sistema de coordenadas. Suponha (x^1, \dots, x^n) e (z^1, \dots, z^n) relacionados por $x^i = x^i(z^1, \dots, z^n)$; se $z^i = z^i(t)$ representar uma curva, teremos cada componente do vetor velocidade no sistema (x^i) dada por $\frac{dx^i}{dt} = \sum_j \frac{\partial x^i}{\partial z^j} \frac{dz^j}{dt}$, e, portanto, chamando de v_x, v_z os vetores velocidade expressos nos sistemas respectivos $(x^i), (z^i)$, temos:

$$|v_x|^2 = \sum_i \left(\frac{dx^i}{dt} \right)^2 = \sum_i \left(\sum_j \frac{\partial x^i}{\partial z^j} v_z^j \right)^2 = \sum_{i,j} g_{ij} v_z^i v_z^j,$$

onde

$$g_{ij} = \sum_k \frac{\partial x^k}{\partial z^i} \frac{\partial x^k}{\partial z^j}. \quad (1.3)$$

Fica também claro que uma mudança no sistema de coordenadas não altera as métricas envolvidas. De fato, se $v = G^{1/2}w$, temos $(v, v)_p = (G^{1/2}w, G^{1/2}w)_p = \langle w, w \rangle_q$, onde $y = G^{1/2}x$, sendo x (resp. y) a representação de p (resp. q).

Chegamos assim à definição de *Métrica Riemanniana* em uma variedade diferenciável M .

É uma correspondência que associa a cada ponto $p \in M$ um produto interno $\langle \cdot, \cdot \rangle_p$ (isto é, uma forma bilinear simétrica, definida positiva) no espaço tangente T_pM , que varia diferenciavelmente no seguinte sentido:

Se $X:U \subset \mathbf{R}^n \rightarrow M$ é um sistema de coordenadas locais em torno de p , com

$$X(x_1, \dots, x_n) = q \in X(U) \quad \text{e} \quad \frac{\partial}{\partial x_i}(q) = dX_q(0, \dots, 1, \dots, 0),$$

então (compare com (3) acima):

$$g_{ij}(x_1, \dots, x_n) := \left\langle \frac{\partial}{\partial x_i}(q), \frac{\partial}{\partial x_j}(q) \right\rangle_q \quad (1.4)$$

é uma função diferenciável.

Observação: As funções $g_{ij}:U \rightarrow \mathbf{R}$, $g_{ij}(x) = \left\langle \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \right\rangle_x$ são chamadas coeficientes da métrica Riemanniana no sistema de coordenadas $X:U \rightarrow M$. Uma variedade diferenciável para a qual se define tal métrica denomina-se uma Variedade Riemanniana.

Exemplo 1.8: $S = \mathbf{R}^n$, $X:\mathbf{R}^n \rightarrow M$, $X(x_1, \dots, x_n) = (x_1, \dots, x_n)$, $T_xM = \mathbf{R}^n$, $dX_x e_i = e_i$; $g_{ij} = \langle e_i, e_j \rangle = \delta_{ij}(x)$.

\mathbf{R}^n tem neste caso as geometrias Riemanniana e Euclidiana coincidentes.

Exemplo 1.9: $M = \{(x, y); y > 0\}$, $T_pM = \mathbf{R}^2$. Para $p = (p_1, p_2) \in M$, definamos $\langle (x_1, y_1), (x_2, y_2) \rangle_p := \frac{1}{p_2} \langle (x_1, y_1), (x_2, y_2) \rangle$, que é, evidentemente uma métrica Riemanniana, explicitada por

$$g_{11}(p) = \langle e_1, e_1 \rangle_p = \frac{1}{p_2}, \quad g_{12}(p) = g_{21}(p) = 0, \quad g_{22}(p) = \frac{1}{p_2}, \quad G = \begin{bmatrix} 1/p_2 & 0 \\ 0 & 1/p_2 \end{bmatrix}.$$

Esta é a métrica (não-euclidiana) de Lobatchevsky (ou Poincaré).

Observação: Os Exemplos 3 e 6 (este, em seu caso particular 7) serão tratados após o estudo de grupos de Lie, estruturas mais ricas que as estudadas até agora, e adaptadas a nossos objetivos.

1.5. Grupo de Lie e Métricas Invariantes por Translação (Escalonamento)

Recordemos que um conjunto S tem uma estrutura de grupo se se definirem as aplicações inversa e produto, respectivamente:

$$x \in S \rightarrow x^{-1} \in S$$

$$x, y \in S \rightarrow xy \in S,$$

verificando:

i) $(xy)z = x(yz)$

ii) $\exists e \in S$ tal que $ex = xe = x$ ($e =$ elemento neutro de S)

iii) $x(x^{-1}) = e$.

Em uma variedade diferenciável com uma estrutura de grupo é natural supor-se que as operações acima sejam diferenciáveis. Tem-se neste caso um grupo de Lie.

Observação: As chamadas translações à esquerda e à direita, definidas respectivamente por

$$L_x: S \rightarrow S, \quad L_x y = xy$$

$$R_x: S \rightarrow S, \quad R_x y = yx$$

são diferenciáveis e possuem respectivas inversas diferenciáveis, ou, conforme nomenclatura usual, são difeomorfismos.

Um grupo simétrico ou Abelianiano é o que verifica a invariância de permutação do produto ($xy = yx$). Em grupos de Lie Abelianos temos $L_x = R_x$.

A partir de agora consideraremos que S é um grupo de Lie Abelianiano.

A fim de construirmos métricas adequadas em S , uma importante propriedade, e que se mostrará útil, é a de que a métrica independa de translações (escalonamento, em nossas aplicações). De modo preciso, dizemos que uma métrica Riemanniana é *invariante por translação à esquerda* se

$$\langle u, v \rangle_y = \langle d(L_x)_y u, d(L_x)_y v \rangle_{L_x y}, \quad (1.5)$$

para todo $x, y \in S$, $u, v \in T_y S$; (5) é a definição de *isometria* para um difeomorfismo, no caso L_x .

Construção de uma métrica invariante por translação

A idéia é referir-se ao elemento neutro, e da seguinte forma: tome um produto $\langle \cdot, \cdot \rangle_e$ qualquer em $T_e S$, e defina

$$\langle u, v \rangle_x := \langle d(L_{x^{-1}})_x u, d(L_{x^{-1}})_x v \rangle_e, \quad x \in S, u, v \in T_x S. \quad (1.6)$$

A verificação de (5) é imediata; além disto L_x depende diferenciavelmente de x , logo (6) define a métrica desejada.

Exemplo 1.10: $S = \text{int}(\mathbf{R}_+^n) = \{x : x > 0\}; T_x S = \mathbf{R}^n$

Estrutura de grupo:

$$xy := (x_1 y_1, \dots, x_n y_n)$$

$$x^{-1} := \left(\frac{1}{x_1}, \dots, \frac{1}{x_n} \right)$$

$$e := (1, \dots, 1)$$

Com esta estrutura, S é um grupo abeliano. A diferenciabilidade das aplicações acima, induzida de \mathbf{R}^n , faz também de S um grupo de Lie.

Observação: Embora aqui apareçam de modo arbitrário, as definições de produto e inversa são naturais, decorrentes da necessidade de definição de e , esta, por sua vez, ponto essencial

na obtenção dos métodos de escalonamento (afim-escala de Dikin [67] e os de Karmarkar [84] e variantes, em geral). Nestes métodos, há um “retorno” ao elemento identidade. Esta explicação é válida para o exemplo a seguir em que o elemento identidade de $\{x \in \mathbf{R}_+^n : e^T x = 1\}$ é e/n .

Vamos agora introduzir uma métrica Riemanniana invariante por translação. Precisamos de:

- i) um produto interno (uma métrica) em $T_e S(\cong \mathbf{R}^n)$; tomaremos o usual $\langle u, v \rangle_e := \sum_i u_i v_i$.

Deve-se observar que esta escolha é natural, sob a perspectiva dos métodos elipsoidais, em Programação Matemática, em que se procura “cobrir” com os maiores elipsóides a região em consideração. No caso, em \mathbf{R}_+^n , o elipsóide centrado em e , $(x - e)^T(x - e) = \text{cte}$ é o que verifica esta propriedade, induzindo portanto a métrica euclidiana usual (veja Figura 6).

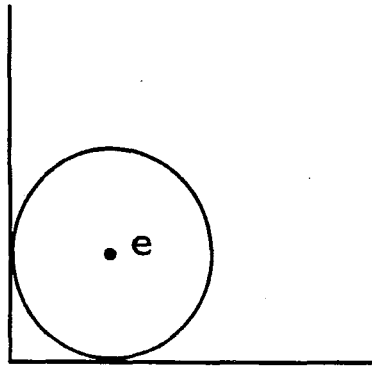


Fig. 6 - E é o maior elipsóide centrado em e , contido em \mathbf{R}_+^2

- ii) calculo de $d(L_{x^{-1}})_y$; temos, sucessivamente:

$$L_x: S \rightarrow S, L_x y = (x_1 y_1, \dots, x_n y_n) = \text{diag}(x_1, \dots, x_n) y$$

$$d(L_x)_y: \mathbf{R}^n \rightarrow \mathbf{R}^n, d(L_x)_y = \text{diag}(x_1, \dots, x_n),$$

$$\text{logo } d(L_{x^{-1}})_y = \text{diag}\left(\frac{1}{x_1}, \dots, \frac{1}{x_n}\right), \forall y \in S.$$

iii) Finalmente,

$$\begin{aligned} \langle u, v \rangle_x &= \langle d(L_{x^{-1}})_x u, d(L_{x^{-1}})_x v \rangle_e \\ &= \left\langle \text{diag} \left(\frac{1}{x_1}, \dots, \frac{1}{x_n} \right) u, \text{diag} \left(\frac{1}{x_1}, \dots, \frac{1}{x_n} \right) v \right\rangle = \sum_i \frac{u_i v_i}{x_i^2}. \end{aligned}$$

e, por conseguinte,

$$G = (g_{ij}(x)) = \left(\frac{\delta_{ij}}{x_i x_j} \right) = D^{-2}, \text{ com } D^{-2} = \text{diag} \left(\frac{1}{x_1^2}, \dots, \frac{1}{x_n^2} \right). \quad (1.7)$$

Trata-se de uma métrica representada por uma matriz diagonal, realizando portanto um escalonamento contínuo do espaço \mathbf{R}_+^n . Em particular, um vetor do espaço tangente, $u \in \mathbf{R}^n$, aplicado em um ponto $x \in \mathbf{R}_+^n$, tem seu comprimento dado por $\ell = \langle u, u \rangle_x^{1/2} = \sqrt{\sum_i u_i^2/x_i^2}$. Os vetores de comprimento constante ℓ se distribuem ao longo do elipsóide $\sum_i u_i^2/x_i^2 = \ell^2$, no espaço tangente a S em x .

Denominamos G de (representação da) métrica afim-escala.

Exemplo 1.11:

$$S = \{x : e^T x = 1, x > 0\}$$

$$T_x S = e^\perp = \{v : e^T v = 0\}$$

Conforme observação feita no exemplo anterior, partindo do centro do simplex S , dado por $e/n = (1/n, \dots, 1/n)$, definâmo-lo como elemento identidade. É razoável definir a estrutura de grupo em S através de

$$\begin{aligned} xy &:= \left(\frac{x_1 y_1}{\sum_i x_i y_i}, \dots, \frac{x_n y_n}{\sum_i x_i y_i} \right) \\ x^{-1} &= \left(\frac{1/x_1}{\sum_i 1/x_i}, \dots, \frac{1/x_n}{\sum_i 1/x_i} \right), \end{aligned}$$

que garantem o fechamento ($xy \in S, x^{-1} \in S$). Estas operações são diferenciáveis, S é portanto um grupo (Abeliano) de Lie.

Obtenção de uma métrica Riemanniana invariante por translações

i) produto interno em $T_{e/n}S$.

Considere a seguinte base em $T_{e/n}S$ (obtida por cálculo direto a partir da expressão do espaço tangente):

$$\beta = \{\alpha_1, \dots, \alpha_{n-1}\}, \quad \alpha_i = u_i - \frac{e}{n} = \left(-\frac{1}{n}, \dots, -\frac{1}{n}, 1 - \frac{1}{n}, -\frac{1}{n}, \dots, -\frac{1}{n} \right),$$

onde a coordenada $1 - 1/n$ ocupa a i -ésima posição e u_i é o i -ésimo vetor unitário. Com $T_{e/n}S$ munido do produto interno euclidiano canônico, definimos

$$g_{ij}(e/n) := (\alpha_i, \alpha_j) = n^2 \left(\delta_{ij} - \frac{1}{n} \right),$$

onde o fator n^2 terá valor puramente estético, conforme ficará claro mais adiante.

ii) Cálculo de $d(L_{x^{-1}})_x$.

Por definição, $L_x: S \rightarrow S$, $L_x y = xy = \frac{x_o y}{x^T y}$, onde denotamos $x_o y = (x_1 y_1, \dots, x_n y_n)$.

Assim

$$\begin{aligned} d(L_x)_{e/n} w &= \frac{(x^T y) x_o w - x_o y (x^T w)}{(x^T y)^2} \Big|_{y=e/n} = \frac{n}{(x^T e)^2} [(x^T e) x_o w - (x^T w) x_o e] \\ &= \frac{n(x^T w)}{x^T e} [xw - xe] \end{aligned}$$

Para $w = \alpha_i$, temos que $d(L_x)_{e/n} \alpha_i$ tem a direção do vetor $x\alpha_i - xe = \frac{x_o \alpha_i}{x^T \alpha_i} - \frac{x_o e}{x^T e}$, que é colinear ao vetor

$$\begin{aligned} &(x^T e) x_o \alpha_i - (x^T \alpha_i) x_o e \\ &= \left(x_o \left(u_i - \frac{e}{n} \right) \right) - \left(x^T \left(u_i - \frac{e}{n} \right) \right) x_o e \\ &= \left(x_i u_i - \frac{1}{n} x \right) - \left(x_i - \frac{1}{n} x^T e \right) x \\ &= x_i u_i - x_i x \end{aligned}$$

e a

$$v_i := u_i - x = (-x_1, \dots, -x_{i-1}, 1 - x_i, -x_{i+1}, \dots, -x_n).$$

Portanto $V := \{v_1, \dots, v_{n-1}\}$ é uma base de $T_x S$. Além disto é fácil ver que

$$d(L_{x^{-1}})_x v_i = \frac{1}{nx_i} \alpha_i.$$

iii) Produto interno em um ponto genérico.

Temos, finalmente, a partir da base V :

$$\begin{aligned} g_{ij}(x) &= (v_i, v_j)_x = (d(L_{x^{-1}})_x v_i, d(L_{x^{-1}})_x v_j)_{e/n} \\ &= \left(\frac{\alpha_i}{nx_i}, \frac{\alpha_j}{nx_j} \right)_{e/n} \\ &= \frac{1}{n^2 x_i x_j} (\alpha_i, \alpha_j)_{e/n} = \frac{1}{n^2 x_i x_j} g_{ij}(e/n) \\ &= \frac{1}{x_i x_j} \left(\delta_{ij} - \frac{1}{n} \right) \end{aligned} \tag{1.8}$$

Observação: Verifica-se uma interessante interpretação para a métrica definida no elemento neutro e/n . Aproveitêmo-nos da base β para definir a matriz $n \times (n-1)$

$$Z := (\alpha_1, \dots, \alpha_{n-1}).$$

Seja agora \tilde{G} uma matriz $n \times n$ representando uma aplicação de \mathbf{R}_+^n em \mathbf{R}_+^n . Propriedades de \tilde{G} restritas a $T_{e/n} S$ (por exemplo, a definição positiva), podem ser expressas através da representação ou “projeção de \tilde{G} ” dada por $Z^T \tilde{G} Z$.

Tome $\tilde{G}(x) = \text{diag} \left(\frac{1}{x_1^2}, \dots, \frac{1}{x_n^2} \right)$, a matriz que define a métrica afim-escala. Por cálculo direto, obtemos:

$$Z^T \tilde{G}(e/n) Z = (n^2 (\alpha_i, \alpha_j)) = G(e/n),$$

isto é, $G(e/n)$ pode ser visto como uma projeção da matriz que representa a métrica afim-escala em e/n , isto é, a métrica no $T_{e/n}(\mathbf{R}_+^n)$ induzida em $T_{e/n}S$.

É necessário observar que, para efeitos de trabalho no espaço \mathbf{R}^n , e por coerência com o resultado acima, definimos $g_{in} = g_{ni} = 0$. Consideramos portanto $v \in T_{e/n}S$ dado pelas $(n-1)$ primeiras coordenadas de $v \in \mathbf{R}^n : (v_1, \dots, v_{n-1}, 1 - \sum_1^{n-1} v_i)$.

Deste modo, a observação feita no exemplo anterior sobre o elipsóide centrado no elemento unitário, aqui e/n , é também válida agora: $(x - \frac{e}{n})^T G(\frac{e}{n})(x - \frac{e}{n}) = \text{cte}$, restrito ao simplex S , corresponde aos “maiores elipsóides” que o cobrem (veja Figura 7).

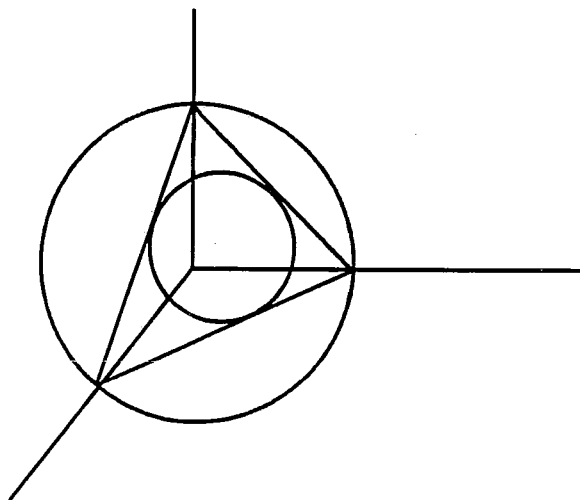


Fig. 7 – O maior (resp. menor) elipsóide inscrito em (resp. circunscrito a) S

1.6. Derivada Covariante, Conexões Afins

Derivada Covariante

Introduzimos os campos de vetores nos espaços tangentes a variedades. Em cada ponto, já o dissemos, eles representam a direção seguida pelas iterações sucessivas do algoritmo considerado em (2). Observados continuamente em S , isto é, através de versão contínua do algoritmo, a trajetória por este seguida terá suas características (curvatura por exemplo) dependentes diretamente das características do campo. Surge assim a necessidade de se ter, para os campos, conceitos correspondentes ao de derivada de funções.

Seja X um campo definido (em cada ponto) na variedade S . Como estaremos interessados em trajetórias em S , consideramos, para uma curva $\alpha: I \subset \mathbf{R} \rightarrow S$, o campo X ao longo da curva, que denominamos $V(t) = X(\alpha(t))$; diz-se que V é induzido por X .

Hipótese: Todos os campos e funções envolvidos serão supostos de classe C^∞ .

Observe que por ser $V(t) \in T_{\alpha(t)}S$, não é obrigatório que a variação de V ao longo da curva, medida em cada ponto por $\frac{dV}{dt}(t)$, também pertença ao espaço tangente. Veremos então que sua projeção neste espaço é equivalente, pelas propriedades que possui, a um processo de diferenciação, denominado covariante. No caso de superfícies, há uma evidente interpretação geométrica para a derivada covariante $\frac{DV}{dt}$ ao longo de $\alpha(t)$: é o campo (tangente) diferenciável, dado por (veja Figura 8, extraída de do Carmo [88]):

$$\frac{DV}{dt} = \frac{dV}{dt} - \left[\frac{dV}{dt} \cdot N(\alpha(t)) \right] N(\alpha(t)),$$

onde N é uma “normal” (e.g., se $S = f^{-1}(c)$, $f: \mathbf{R}^n \rightarrow \mathbf{R}$, $N = \pm \frac{\nabla f(\alpha(t))}{\|\nabla f(\alpha(t))\|}$).

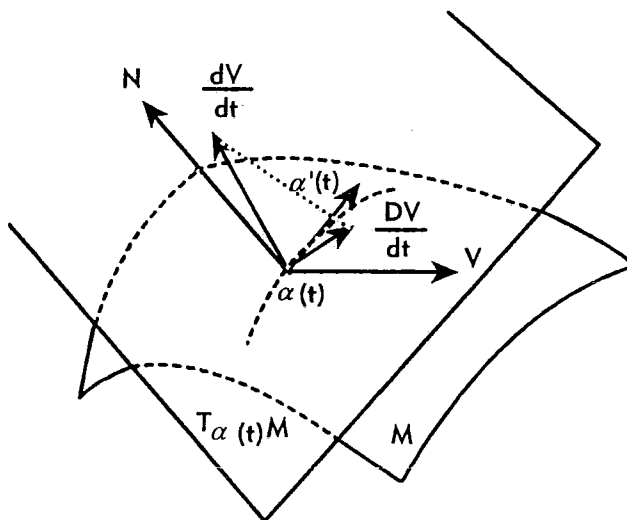


Fig. 8 – As componentes tangente e normal da derivada de V

Esta expressão verifica as seguintes propriedades, que serão usadas como definição de Derivada Covariante de V ao longo de α :

- $\frac{D}{dt}(V + W) = \frac{DV}{dt} + \frac{DW}{dt}$, para qualquer outro campo W , diferenciável ao longo de α ;
- $\frac{D}{dt}(fV) = f \frac{DV}{dt} + \frac{df}{dt}V$, para qualquer função diferenciável $f: S \rightarrow \mathbf{R}$.

Conexões Afins

Reportando-nos à interpretação geométrica acima, e lembrando que V é induzido por X em α , poderíamos “ver” $\frac{DV}{dt}$ como uma medida da variação de X segundo as direções dadas pelo campo $Y(t) := \alpha'(t)$ dos vetores tangentes a $\alpha(t)$. Isto é a chamada conexão afim, definida em uma variedade diferenciável S , denotada por $\nabla_Y X$, que, neste caso, se escreve

$$\nabla_Y X = \nabla_{\frac{d\alpha}{dt}} X = \frac{DV}{dt} \quad (1.9)$$

De um modo geral, para campos em $\Omega := \mathcal{X}(S)$ e funções reais em $D = C^\infty(S)$, uma conexão afim é uma aplicação que verifica:

$$\nabla: \Omega \times \Omega \rightarrow \Omega,$$

$$(X, Y) \rightarrow \nabla_X Y$$

$$\text{i) } \nabla_{(fX+gY)}Z = f\nabla_XZ + g\nabla_YZ$$

$$\text{ii) } \nabla_X(Y + Z) = \nabla_XY + \nabla_XZ$$

$$\text{iii) } \nabla_XfY = f\nabla_XY + X(f)Y,$$

onde $X, Y, Z \in \Omega$, e $f, g \in D$.

Observação: Demonstra-se que a aplicação que associa V , campo ao longo de α , a DV/dt , derivada covariante ao longo de α , tal que DV/dt verifica a), b) e (9), é única.

Expressão de conexões relativamente a coordenadas locais

Suponha que os campos $X, Y \in \Omega$ sejam representados em uma certa vizinhança local $\mathcal{X}: U \subset \mathbb{R}^n \rightarrow S$, de algum ponto p , por

$$X = \sum_{i=1}^n x_i \frac{\partial}{\partial x_i}, \quad Y = \sum_{i=1}^n y_i \frac{\partial}{\partial x_i}$$

($\partial/\partial x_i$ representam os vetores da base do sistema de coordenadas locais).

Temos, segundo as propriedades de definição de conexão afim:

$$\begin{aligned} \nabla_X Y &= \nabla_{\sum x_i \frac{\partial}{\partial x_i}} \left[\sum_j y_j \frac{\partial}{\partial x_j} \right] = \sum_i x_i \left[\nabla_{\frac{\partial}{\partial x_i}} \left[\sum_j y_j \frac{\partial}{\partial x_j} \right] \right] = \\ &= \sum_i x_i \left[\sum_j \left[y_j \nabla_{\frac{\partial}{\partial x_i}} \frac{\partial}{\partial x_j} \right] \right] + \sum_i x_i \left[\sum_j \left[\frac{\partial}{\partial x_i} (y_j) \frac{\partial}{\partial x_j} \right] \right]. \end{aligned}$$

Observe que $\frac{\nabla_{\partial}}{\partial x_i} \frac{\partial}{\partial x_j} \in \Omega$, podendo portanto ser também representado através da base local, isto é:

$$\frac{\nabla_{\partial}}{\partial x_i} \frac{\partial}{\partial x_j} = \sum_{k=1}^n \Gamma_{ij}^k \frac{\partial}{\partial x_k}, \quad (1.10)$$

que, substituído na equação acima, fornece:

$$\nabla_X Y = \sum_k \left[\sum_{i,j} x_i y_j \Gamma_{ij}^k + \sum_i x_i \frac{\partial}{\partial x_i} (y_k) \right] \frac{\partial}{\partial x_k}.$$

Definição 1.1: Símbolos de Christoffel, ou coeficientes da conexão afim ∇ em U , são as funções (diferenciáveis):

$$\Gamma_{ij}^k: U \subset S \rightarrow \mathbf{R},$$

definidas por (10).

Expressão da derivada covariante em termos de coordenadas locais e dos símbolos de Christoffel

Seja, como acima, $\mathcal{X}: U \rightarrow S$ um sistema de coordenadas locais em torno de $P \in S$.

Considere a curva $\alpha: I \rightarrow \mathcal{X}(U)$, diferenciável, expressa por $\alpha(t) = (\alpha_1(t), \dots, \alpha_n(t))$, e o campo ao longo de α definido por $V(t) = \sum_k v_k(t) \frac{\partial}{\partial x_k}(t)$. O campo tangente a α é

$$\alpha' = \frac{d\alpha}{dt} = \sum_i \frac{d\alpha_j}{dt} \frac{\partial}{\partial x_i}.$$

Portanto, utilizando propriedades da derivada covariante, obtemos

$$\begin{aligned} \frac{DV}{dt} &= \frac{D}{dt} \left(\sum_k v_k \frac{\partial}{\partial x_k} \right) = \sum_k \frac{D}{dt} \left(v_k \frac{\partial}{\partial x_k} \right) \\ &= \sum_k \left(\frac{dv_k}{dt} \frac{\partial}{\partial x_k} + v_k \frac{D}{dt} \left(\frac{\partial}{\partial x_k} \right) \right) \end{aligned} \tag{1.11}$$

Por outro lado,

$$\begin{aligned} \frac{D}{dt} \left(\frac{\partial}{\partial x_k} \right) &= \nabla_{\frac{d\alpha}{dt}} \frac{\partial}{\partial x_k} = \nabla_{\sum_i \frac{d\alpha_j}{dt} \frac{\partial}{\partial x_i}} \frac{\partial}{\partial x_k} \\ &= \sum_i \frac{d\alpha_j}{dt} \nabla_{\frac{\partial}{\partial x_i}} \frac{\partial}{\partial x_k}. \end{aligned}$$

A última expressão é a que define, segundo (10), os símbolos de Christoffel Γ_{ik}^j , que,

substituído em (11), fornece, para DV/dt :

$$\begin{aligned} \frac{DV}{dt} &= \sum_k \left(\frac{dv_k}{dt} \frac{\partial}{\partial x_k} + v_k \sum_{i,j} \frac{d\alpha_i}{dt} \Gamma_{ik}^i \frac{\partial}{\partial x_j} \right) \\ &= \sum_k \left(\frac{dv_k}{dt} + \sum_{i,j} v_j \frac{d\alpha_i}{dt} \Gamma_{ij}^k \right) \frac{\partial}{\partial x_k} \end{aligned} \tag{1.12}$$

1.7. Campos Paralelos e Geodésicas

Em \mathbf{R}^n munido com a métrica Euclidiana a geodésica é o segmento de reta entre dois pontos p e q quaisquer que caracteriza a trajetória de menor comprimento que os liga. Se $\alpha: [a, b] \rightarrow \mathbf{R}^n$, $\alpha(a) = p$, $\alpha(b) = q$ é uma curva diferenciável passando por p e q , sendo o campo tangente $\frac{d\alpha}{dt} = \left(\frac{d\alpha_1}{dt}, \dots, \frac{d\alpha_n}{dt} \right)$ associado fisicamente à velocidade, temos a aceleração $\frac{d}{dt} \left(\frac{d\alpha}{dt} \right) = \left(\frac{d^2\alpha_1}{dt^2}, \dots, \frac{d^2\alpha_n}{dt^2} \right)$ em cada ponto $\alpha(t)$ definindo a propriedade de α ser geodésica dada por: $\frac{d}{dt} \left(\frac{d\alpha}{dt} \right) = 0$.

A extensão desta noção às variedades exigirá apenas que a componente tangencial da derivada seja nula.

Geodésica

Uma curva parametrizada $\alpha: I \rightarrow S$ é uma geodésica se o campo tangente $\alpha'(t)$ de α verifica

$$\frac{D}{dt} \left(\frac{d\alpha}{dt} \right) = 0.$$

Campos Paralelos

De um modo geral, dados uma variedade S , uma conexão afim ∇ e um campo V ao longo de uma curva diferenciável $\alpha: I \rightarrow S$, V é denominado paralelo (compare com \mathbf{R}^n euclidiano) se $\frac{DV}{dt} = 0$, para todo $t \in I$.

Assim, se α for uma geodésica, α' é paralelo.

Equações de Geodésica

Dada a expressão (12), um campo V paralelo é determinado pela equação

$$\sum_k \left(\frac{dv_k}{dt} + \sum_{i,j} v_j \frac{d\alpha_i}{dt} \Gamma_{ij}^k \right) \frac{\partial}{\partial x_k} = 0,$$

ou, equivalentemente, pelo sistema de n equações diferenciais em v_k (os elementos $\frac{\partial}{\partial x_k}$ são linearmente independentes):

$$\frac{dv_k}{dt} + \sum_{i,j} v_j \frac{d\alpha_i}{dt} \Gamma_{ij}^k = 0, \quad k = 1, \dots, n.$$

Quando se trata de uma geodésica $\alpha(t) = (x_1(t), \dots, x_n(t))$, tem-se $v_i = \frac{dx_i}{dt}$, e o último sistema se transforma em

$$\frac{d}{dt} \left(\frac{dx_k}{dt} \right) + \sum_{i,j} \frac{dx_j}{dt} \frac{dx_i}{dt} \Gamma_{ij}^k = 0, \quad k = 1, \dots, n$$

ou

$$\frac{d^2 x_k}{dt^2} + \sum_{i,j} \Gamma_{ij}^k \frac{dx_i}{dt} \frac{dx_j}{dt} = 0, \quad k = 1, \dots, n, \quad (1.13)$$

sistema de n equações diferenciais de 2ª ordem, que possui solução única em algum intervalo $I = [a, b]$, verificando $x(a) = \alpha(a)$ e $\frac{dx}{dt}(a) = \frac{d\alpha}{dt}(a)$, dados (resultante da teoria de Equações Diferenciais Ordinárias).

1.8. Conexões Afins em Variedades Riemannianas

Um estudo detalhado dos símbolos de Christoffel e derivada covariante em superfícies de \mathbf{R}^3 (veja do Carmo [76]), definidas por exemplo por $f^{-1}(c)$, para $f: \mathbf{R}^3 \rightarrow \mathbf{R}$, leva a

expressões dependentes de $\frac{\partial^2 f}{\partial x_i \partial x_j}$, de que se conclui que aqueles símbolos são simétricos em ij : $\Gamma_{ij}^k = \Gamma_{ji}^k$, e, conseqüentemente, a conexão afim expressa por meio destes coeficientes também é simétrica (veja (10)): $\nabla_{\partial/\partial x_i} \partial/\partial x_j = \nabla_{\partial/\partial x_j} \partial/\partial x_i$, para cada sistema de coordenadas locais.

A extensão dos conceitos vistos nos números anteriores às variedades Riemannianas, em que se define uma métrica através de uma matriz simétrica e de um produto escalar (simétrico), torna natural a exigência de

Conexões afins simétricas: $\nabla_{\frac{\partial}{\partial x_i}} \frac{\partial}{\partial x_j} = \nabla_{\frac{\partial}{\partial x_j}} \frac{\partial}{\partial x_i}$.

Será também necessária a

Definição 1.2: Seja M uma variedade diferenciável com uma conexão afim ∇ e uma métrica Riemanniana $\langle \cdot, \cdot \rangle$. Diz-se que ∇ é *compatível* com a métrica $\langle \cdot, \cdot \rangle$ se e só se para todo par de campos de vetores V e W ao longo da curva diferenciável $\alpha: I \rightarrow S$ se tem

$$\frac{d}{dt} \langle V, W \rangle = \left\langle \frac{DV}{dt}, W \right\rangle + \left\langle V, \frac{DW}{dt} \right\rangle, \quad (1.14)$$

que é a regra usual de derivada de produto.

Observação: Se $\alpha(t) = (x_1(t), \dots, x_n(t))$ for geodésica, (14) leva a

$$\frac{d}{dt} \left\langle \frac{d\alpha}{dt}, \frac{d\alpha}{dt} \right\rangle = \left\langle \frac{D}{dt} \frac{d\alpha}{dt}, \frac{d\alpha}{dt} \right\rangle + \left\langle \frac{d\alpha}{dt}, \frac{D}{dt} \frac{d\alpha}{dt} \right\rangle = 0,$$

e, conseqüentemente, $\left\langle \frac{d\alpha}{dt}, \frac{d\alpha}{dt} \right\rangle$ é constante.

O teorema a seguir é o instrumento teórico que garante a existência e unicidade de conexão adequada à métrica Riemanniana:

Teorema 1.1 (Levi-Civita): Dada uma variedade Riemanniana M , existe uma única conexão afim ∇ em M satisfazendo as condições:

- a) ∇ é simétrica;

b) ∇ é compatível com a métrica Riemanniana.

(Esta conexão é denominada *Conexão Riemanniana*).

Prova: do Carmo [88].

Relação entre a métrica Riemanniana e os símbolos de Christoffel

$G(x) = (g_{ij}(x))$ admite uma inversa $G^{-1}(x) = (g^{ij}(x))$; mostra-se que

$$\Gamma_{ij}^m = \frac{1}{2} \sum_k \left\{ \frac{\partial}{\partial x_i} g_{ik} + \frac{\partial}{\partial x_j} g_{ki} - \frac{\partial}{\partial x_k} g_{ij} \right\} g^{km} \quad (1.15)$$

Exemplo 1.12: $M = \mathbf{R}^n$, com a métrica usual, $ds = \sum_i dx_i dx_i$. Neste caso,

$$g_{ij}(x) = \delta_{ij}, \text{ isto é, } G(x) = I,$$

$$g^{ij}(x) = \delta_{ij}, \Gamma_{ij}^k = 0, \text{ para todo } i, j, k = 1, \dots, n.$$

Assim se $V(t) = \sum_k v_k \frac{\partial}{\partial x_k}$ é um campo ao longo de uma curva α , temos $\frac{DV}{dt} =$

$\sum_k \frac{dv_k}{dt} \frac{\partial}{\partial x_k}$, que é a própria derivada usual. Além disto, as equações das geodésicas em \mathbf{R}^n

(veja (13)) são

$$\frac{d^2 x_k}{dt^2} = 0, \quad k = 1, \dots, n,$$

cujas soluções são da forma

$$x_k(t) = a_k t + b_k.$$

$$\alpha: \mathbf{R} \rightarrow M, \quad \alpha(t) = (a_1 t + b_1, \dots, a_n t + b_n);$$

Exemplo 1.13: $M = \text{int } \mathbf{R}_+^n$, com a métrica já definida $ds^2 = \sum_i \frac{1}{x_i^2} dx_i dx_i$, isto é

$$g_{ij}(x) = \frac{\delta_{ij}}{x_i x_j}, \quad \text{ou } G(x) = \text{diag} \left(\frac{1}{x_1^2}, \dots, \frac{1}{x_n^2} \right),$$

de que se obtém

$$G^{-1}(x) = \text{diag}(x_1^2, \dots, x_n^2) \text{ e } g^{ij}(x) = x_i x_j \delta_{ij}.$$

Cálculo dos símbolos de Cristoffel

Repitamos a fórmula (15):

$$\Gamma_{ij}^m = \frac{1}{2} \sum_k \left\{ \frac{\partial}{\partial x_i} g_{jk} + \frac{\partial}{\partial x_j} g_{ki} - \frac{\partial}{\partial x_k} g_{ij} \right\} g^{km}$$

i) $i \neq j$, fixados: se $k \neq m$, $g^{km} = 0$, e então

$$\Gamma_{ij}^m = \frac{1}{2} \left\{ \frac{\partial}{\partial x_i} g_{jm} + \frac{\partial}{\partial x_j} g_{mi} \right\} g^{mm}.$$

É fácil ver que se $m \neq i$, $m \neq j$ ou $m \neq i, j$, se tem

$$\Gamma_{ij}^m = 0.$$

ii) $i = j$ fixados; também aqui só interessa o caso em que $k = m$:

$$\Gamma_{ii}^m = \frac{1}{2} \left\{ \frac{\partial}{\partial x_i} g_{im} + \frac{\partial}{\partial x_i} g_{mi} - \frac{\partial}{\partial x_m} g_{ii} \right\} g^{mm};$$

se $m \neq i$, $\Gamma_{ii}^m = 0$, e, finalmente, se $m = i$,

$$\Gamma_{ii}^i = \frac{1}{2} \left(\frac{\partial}{\partial x_i} g_{ii} \right) g^{ii} = \frac{1}{2} \left\{ -\frac{2}{x_i^3} \right\} x_i^2 = -\frac{1}{x_i}.$$

Resumindo:

$$\Gamma_{ij}^m = -\frac{1}{x_i} \delta_{im} \delta_{ij}, \quad (1.16)$$

e, segundo (10),

$$\nabla_{\frac{\partial}{\partial x_i}} \frac{\partial}{\partial x_j} = -\frac{\delta_{ij}}{x_i} \frac{\partial}{\partial x_i}.$$

Equações Geodésicas

Se $V(t) = \sum_m v_m \frac{\partial}{\partial x_m}$ é um campo ao longo de α , a expressão (12) implica em

$$\begin{aligned} \frac{DV}{dt} &= \sum_m \left(\frac{dv_m}{dt} + \sum_{i,j} v_j \frac{d\alpha_i}{dt} \Gamma_{ij}^m \right) \frac{\partial}{\partial x_m} \\ &= \sum_m \left(\frac{dv_m}{dt} + \sum_i v_i \frac{d\alpha_i}{dt} \Gamma_{ii}^m \right) \frac{\partial}{\partial x_m} \\ &= \sum_m \left\{ \frac{dv_m}{dt} + v_m \frac{d\alpha_m}{dt} \Gamma_{mm}^m \right\} \frac{\partial}{\partial x_m} \\ &= \sum_m \left\{ \frac{dv_m}{dt} - \frac{v_m}{\alpha_m} \frac{d\alpha_m}{dt} \right\} \frac{\partial}{\partial x_m}. \end{aligned}$$

Se $\alpha(t) = (x_1(t), \dots, x_n(t))$ for geodésica, o campo $V(t) = \frac{d\alpha}{dt}(t)$ fornece então o sistema

$$\frac{d^2 x_m}{dt^2} - \frac{1}{x_m} \left(\frac{dx_m}{dt} \right)^2 = 0, \quad m = 1, \dots, n.$$

Estas equações são independentes, cada uma podendo ser reescrita como

$$0 = \frac{1}{x_m} \frac{d^2 x_m}{dt^2} - \frac{1}{x_m^2} \left(\frac{dx_m}{dt} \right)^2 = \frac{d}{dt} \left(\frac{1}{x_m} \frac{dx_m}{dt} \right).$$

Dadas as condições iniciais, $x(0) = b > 0$ e $\dot{x}(0) = a \in \mathbf{R}^n$ obtemos

$$\begin{aligned} \frac{1}{x_m} \frac{dx_m}{dt} &= \frac{a_m}{b_m} \Rightarrow \log x_m(s) \Big|_{s=0}^{s=t} = \frac{a_m}{b_m} t, \\ \frac{x_m(t)}{b_m} &= \exp \left(\frac{a_m t}{b_m} \right) \Rightarrow x_m(t) = b_m \exp \left(\frac{a_m}{b_m} t \right). \end{aligned}$$

Assim, a geodésica de S , com condições iniciais $\alpha(0) = b \in \dot{\alpha}(0) = a$, é a curva $\alpha: \mathbf{R} \rightarrow S$,

$$\alpha(t) = \left(b_1 \exp \left(\frac{a_1}{b_1} t \right), \dots, b_n \exp \left(\frac{a_n}{b_n} t \right) \right) \quad (1.17)$$

Exemplo 1.14: $S = \{x \in \mathbf{R}^n; e^T x = 1, x > 0\}$, com a métrica já definida $g_{ij}(x) = \frac{1}{x_i x_j} [\delta_{ij} - \frac{1}{n}]$, cujos símbolos de Cristoffel são dados por

$$\Gamma_{ij}^m(x) = -\delta_{ij} \left[\frac{\delta_{im}}{x_m} - \frac{x_m}{x_i} \right], \quad (1.18)$$

veja Karmarkar [90], pág. 58.

Substituindo (18) na equação das geodésicas (13), obtemos:

$$\frac{d^2 x_m}{dt^2} + \sum_{i=1}^n \left[-\frac{\delta_{im}}{x_m} + \frac{x_m}{x_i} \right] \left(\frac{dx_i}{dt} \right)^2 = 0, \quad m = 1, \dots, n \quad (1.19)$$

Deduziremos a seguir a solução para estas equações.

Dividindo cada uma das equações por x_m , obtemos:

$$\frac{d}{dt} \left(\frac{1}{x_m} \frac{dx_m}{dt} \right) = - \sum_{i=1}^n \frac{1}{x_i} \left(\frac{dx_i}{dt} \right)^2,$$

somando em m ,

$$\frac{d}{dt} \left(\frac{1}{n} \sum_m \frac{1}{x_m} \frac{dx_m}{dt} \right) = - \sum_{i=1}^n \frac{1}{x_i} \left(\frac{dx_i}{dt} \right)^2.$$

Subtraindo da anterior a última igualdade, obtemos agora,

$$\frac{d}{dt} \left(\frac{1}{x_n} \frac{dx_m}{dt} - \frac{1}{n} \sum_{k=1}^n \frac{1}{x_k} \frac{dx_k}{dt} \right) = 0, \quad m = 1, \dots, n \quad (1.20)$$

Fixemos as condições iniciais:

$$x(0) = a \in S \in \dot{x}(0) = b \in T_a S = \{y \in \mathbf{R}^n; e^T y = 0\}.$$

De (20) temos que para cada m ,

$$\frac{1}{x_m} \frac{dx_m}{dt} - \frac{1}{n} \sum_{k=1}^n \frac{1}{x_k} \frac{dx_k}{dt} = \lambda_m,$$

com

$$\lambda_m = \frac{b_m}{a_m} - \frac{1}{n} \sum_{k=1}^n \frac{b_k}{a_k}$$

$$\lambda_m - \lambda_1 = \frac{b_m}{a_m} - \frac{b_1}{a_1} = \frac{1}{x_m} \left(\frac{dx_m}{dt} \right) - \frac{1}{x_1} \left(\frac{dx_1}{dt} \right).$$

Portanto,

$$\frac{x_m(t)}{x_1(t)} = \frac{a_m}{a_1} \exp \left[\left(\frac{b_m}{a_m} - \frac{b_1}{a_1} \right) t \right].$$

Da igualdade acima e usando o fato de que $x(t) \in S$, obtemos:

$$x_m(t) = a_m \exp \left[\left(\frac{b_m}{a_m} - \frac{b_1}{a_1} \right) t \right] / \left(\sum_{k=1}^n a_k \exp \left[\left(\frac{b_k}{a_k} - \frac{b_1}{a_1} \right) t \right] \right), \quad (1.21)$$

$m = 1, \dots, n$.

1.9. Curvaturas

Curvatura Riemanniana

Nesta seção apresentaremos a definição algébrica de curvatura e estudaremos suas propriedades. O que se perde por não abordarmos o significado geométrico é compensado pela facilidade de manipulação algébrica deste conceito.

Definição 1.3. Seja M uma variedade Riemanniana com a conexão de Levi-Civita ∇ . A função $R: \Omega^3 \rightarrow \Omega$ dada por

$$R(X, Y)Z = \nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z + \nabla_{[X, Y]} Z$$

é chamada a *curvatura Riemanniana* de M . Nesta definição aparece o símbolo do colchete dos campos X e Y : $[X, Y]$. Como nos limitaremos à apresentação de resultados sem demonstração, basta, aqui, dizermos que $[X, Y] = YX - XY \in \Omega$.

Observação:

- 1) se $M = \mathbf{R}^n$, é imediato verificar que $R(X, Y)Z = 0$, $\forall X, Y$ e $Z \in \Omega$;
- 2) se (U, x) é um sistema de coordenadas em torno de $p \in M$ e $\left\{ \frac{\partial}{\partial x_i}; i = 1, \dots, n \right\}$ é uma base de $T_p M$, então

$$R\left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}\right) \frac{\partial}{\partial x_k} = \left(\nabla_{\frac{\partial}{\partial x_j}} \nabla_{\frac{\partial}{\partial x_i}} - \nabla_{\frac{\partial}{\partial x_i}} \nabla_{\frac{\partial}{\partial x_j}} \right) \frac{\partial}{\partial x_k}$$

pois, $\left[\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \right] = 0$. Assim diríamos que, a curvatura mede a não comutatividade da derivada covariante.

Teorema 1.2. A curvatura R de uma variedade Riemanniana é tri-linear.

Prova: do Carmo [88], pág. 90. ■

Vamos agora obter a expressão da curvatura em coordenadas locais.

Teorema 1.3. Seja (U, x) um sistema de coordenadas em torno de $p \in M$ e $\left\{ \frac{\partial}{\partial x_i}, i = 1, \dots, n \right\}$ uma base de $T_p M$. Neste sistema de coordenadas

$$R\left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}\right) \frac{\partial}{\partial x_k} = \sum_{\ell} R_{ijk}^{\ell} \frac{\partial}{\partial x_{\ell}}$$

onde as componentes R_{ijk}^{ℓ} de R são dadas por

$$R_{ijk}^{\ell} = \frac{\partial}{\partial x_j} \Gamma_{ik}^{\ell} - \frac{\partial}{\partial x_i} \Gamma_{jk}^{\ell} + \sum_s \Gamma_{ik}^s \Gamma_{js}^{\ell} - \sum_s \Gamma_{jk}^s \Gamma_{is}^{\ell}.$$

Prova: do Carmo [88], pág. 92. ■

Observação: Se nas coordenadas (U, x) , escrevemos

$$X = \sum_i u^i \frac{\partial}{\partial x_i}, \quad Y = \sum_j v^j \frac{\partial}{\partial x_j} \quad \text{e} \quad Z = \sum_k w^k \frac{\partial}{\partial x_k}$$

então, pela linearidade de R , temos

$$R(X, Y)Z = \sum_{i,j,k,\ell} R_{ijk}^\ell u^i v^j w^k \frac{\partial}{\partial x_\ell}.$$

Assim os coeficientes R_{ijk}^ℓ da curvatura só dependem dos coeficientes Γ_{ij}^k (símbolos de Christoffel) da conexão e das suas derivadas.

Curvatura Seccional

Passemos agora ao conceito de curvatura seccional (com valor real), que, embora bem mais simples que R , em um certo sentido a determina.

Seja σ um subespaço bi-dimensional do plano tangente $T_p M$ – defina a forma quadrática $Q: \sigma \rightarrow \mathbf{R}$ por

$$Q(x, y) = \langle x, x \rangle \langle y, y \rangle - \langle x, y \rangle^2.$$

Observação: Geometricamente $|Q(x, y)|$ representa a área do paralelogramo determinado por x e y .

Definição 1.4. $R(x, y)z := R(X, Y)Z(p)$, para qualquer conjunto de campos X, Y e $Z \in \Omega$ tais que, $X(p) = x, Y(p) = y$ e $Z(p) = z$.

Teorema 1.4. Sejam $\sigma \subset T_p M$ um subespaço bi-dimensional e $x, y \in \sigma$, dois vetores linearmente independentes. Então

$$K(x, y) := \frac{\langle R(x, y)x, y \rangle}{Q(x, y)}$$

não depende da escolha dos vetores $x, y \in \sigma$.

Prova: do Carmo [88], pág. 94. ■

Definição 1.5. Dados um ponto $p \in M$ e um subespaço bi-dimensional $\sigma \subset T_p M$, o número real $K(x, y) = k(\sigma)$, onde $\{x, y\}$ é uma base qualquer de σ , é chamado *curvatura seccional* de σ em p .

Observação: Segue da definição que a curvatura seccional k de M é uma função de valor real, definida no conjunto dos subespaços bi-dimensionais tangentes a M .

Por definição, R determina k ; veremos agora em que sentido k determina R .

Suponhamos que $\dim(T_p M) \geq 2$ e consideremos as seguintes aplicações tri-lineares

$$R^1: T_p M^3 \rightarrow T_p M \quad \text{e} \quad R^2: T_p M^3 \rightarrow T_p M,$$

satisfazendo as seguintes condições:

- (a) $\langle R^i(x, y)z, t \rangle = -\langle R^i(y, x)z, t \rangle;$
- (b) $\langle R^i(x, y)z, t \rangle = -\langle R^i(x, y)t, z \rangle;$
- (c) $R^i(x, y)z + R^i(y, x)x + R^i(z, x)y = 0;$
- (d) $\langle R^i(x, y)z, t \rangle = \langle R^i(z, t)x, y \rangle;$

para $i = 1, 2$.

Observamos que a curvatura Riemanniana R de M , satisfaz (a), (b), (c) e (d).

Sejam agora x, y dois vetores linearmente independentes de $T_p M$; definamos

$$k^1(\sigma) := \frac{\langle R^1(x, y), x, y \rangle}{Q(x, y)}, \quad k^2(\sigma) := \frac{\langle R^2(x, y), x, y \rangle}{Q(x, y)}$$

onde σ é o subespaço bi-dimensional gerado por x e y . Se para todo $\sigma \subset T_p M$, $k^1(\sigma) = k^2(\sigma)$, então $R^1 = R^2$.

Exemplo 1.15: Seja $S = \{x \in \mathbf{R}^n; x > 0\}$ com a estrutura de variedade vista nos números anteriores. Como vimos em (16) seus símbolos de Christoffel são dados por

$$\Gamma_{ij}^\ell = -\frac{1}{x_i} \delta_{i\ell} \delta_{ij}$$

e se tem

$$\nabla_{\frac{\partial}{\partial x_i}} \frac{\partial}{\partial x_k} = - \left(\frac{1}{x_i} \delta_{ik} \right) \frac{\partial}{\partial x_i}.$$

Deste modo

$$\nabla_{\frac{\partial}{\partial x_j}} \nabla_{\frac{\partial}{\partial x_i}} \frac{\partial}{\partial x_k} = -\delta_{ik} \nabla_{\frac{\partial}{\partial x_j}} \left(\frac{1}{x_i} \right) \frac{\partial}{\partial x_i}.$$

Temos a analisar os seguintes casos:

- (a) $i \neq k$, que determina $\nabla_{\frac{\partial}{\partial x_j}} \nabla_{\frac{\partial}{\partial x_i}} \frac{\partial}{\partial x_k} = 0, \forall i$.
- (b) $i = k$ e $j \neq k$, e então

$$\nabla_{\frac{\partial}{\partial x_j}} \nabla_{\frac{\partial}{\partial x_i}} \frac{\partial}{\partial x_k} = \left(\frac{1}{x_k} \right) \left(\frac{1}{x_j} \delta_{jk} \right) \frac{\partial}{\partial x_j} = 0.$$

Do mesmo modo

$$\nabla_{\frac{\partial}{\partial x_i}} \nabla_{\frac{\partial}{\partial x_j}} \frac{\partial}{\partial x_k} = 0, \text{ se } i \neq k \text{ ou se } i = k \text{ e } j \neq k.$$

Assim,

$$R \left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \right) \frac{\partial}{\partial x_k} = \nabla_{\frac{\partial}{\partial x_j}} \nabla_{\frac{\partial}{\partial x_i}} \frac{\partial}{\partial x_k} - \nabla_{\frac{\partial}{\partial x_i}} \nabla_{\frac{\partial}{\partial x_j}} \frac{\partial}{\partial x_k} = 0$$

$\forall i, j, k$.

Portanto $R(X, Y)Z = 0, \forall X, Y, Z \in x(S)$, isto é, a curvatura R de S é nula.

Exemplo 1.16: $S = \{x \in \mathbb{R}^n; e^T x = 1, x > 0\}$, com a métrica $g_{ij}(x) = \frac{1}{x_i x_j} [\delta_{ij} - \frac{1}{n}]$.

Usando os símbolos de Christoffel dados por (20), pode-se mostrar que a curvatura R de S é nula.

1.10. Subvariedades

O objetivo deste número é o estabelecimento de alguns conceitos e propriedades na passagem de variedades a subvariedades.

Sejam M^n e N^n variedades diferenciáveis. Uma aplicação diferenciável $\varphi: M \rightarrow N$ é uma imersão se $d\varphi_p: T_p M \rightarrow T_{\varphi(p)} N$ é injetiva para todo $p \in M$. Se, além disso, φ é um homeomorfismo sobre $\varphi(M) \subset N$, onde $\varphi(M)$ tem a topologia induzida por N , diz-se que φ é um mergulho. Se $M \subset N$ e a inclusão

$$i: M \rightarrow N \text{ é um mergulho}$$

diz-se que M é uma subvariedade de N .

Na maior parte das questões locais de geometria é indiferente tratar com imersões ou mergulhos. Isto provém da seguinte proposição que mostra ser toda imersão, localmente, um mergulho.

Teorema 1.5. Seja $\varphi: M_1^n \rightarrow M_2^m$, $n \leq m$, uma imersão da variedade M_1 na variedade M_2 . Para todo $p \in M_1$, existe uma vizinhança $V \subset M_1$ de p tal que a restrição $\varphi: V \rightarrow M_2$ é um mergulho.

Prova: Veja do Carmo [88], pág. 13. ■

Métrica Induzida

Estamos interessados no caso em que N tem uma métrica Riemanniana $\langle \cdot, \cdot \rangle$, isto é, N é uma variedade Riemanniana, e consideramos em M a métrica induzida de N , isto é,

$$\begin{aligned} p \in M, \quad v, w \in T_p M \\ \langle v, w \rangle'_P = \langle v, w \rangle_p. \end{aligned} \tag{1.20}$$

Tem-se que M , munida desta métrica é uma variedade Riemanniana, chamada subvariedade Riemanniana de N .

Para cada $p \in M$, $T_p N$ pode ser visto com o produto interno $\langle \cdot \rangle_p$, e o subespaço $T_p M$ tem um complemento ortogonal, denotado por $(T_p M)^\perp \subset T_p N$; podemos então usar a decomposição

$$T_p N = T_p M \oplus (T_p M)^\perp$$

para definir duas projeções

$$\begin{aligned} T: T_p N &\rightarrow T_p M \quad (\text{projeção tangencial}) \\ \perp: T_p N &\rightarrow (T_p M)^\perp \quad (\text{projeção normal}) \end{aligned} \tag{1.22}$$

tais que

$$\begin{aligned} \forall v \in T_p N \\ v = Tv + \perp v. \end{aligned}$$

Exemplo 1.17: Seja A matriz $m \times n$, posto de $(A) = m < n$, e seja $N = \text{int}(\mathbf{R}_+^n)$ com métrica Riemanniana $G(x) = D^{-2}$, onde $D = \text{diag}(x_1, \dots, x_n)$. Definimos

$$M = \{x \in N \mid Ax = b, b \in \mathbf{R}^m\}.$$

Da hipótese de posto $(A) = m$, temos que M é subvariedade de N e podemos considerar M com a métrica induzida de N . Estamos interessados em determinar as projeções; para isso, observemos que para

$$x \in M, \quad T_x M = \{v \in \mathbf{R}^n \mid Av = 0\}$$

e a matriz adjunta de A , denotada por $A^*(x)$, é dada por

$$A^*(x) = G^{-1}(x)A^T = D^2 A^T.$$

É fácil ver que $v \in T_x N$, $v = v_T + v_N$ verificando

$$Av_T = 0$$

$$v_N = A^*w, \quad w \in \mathbf{R}^m$$

Assim,

$$v = v_T + D^2 A^T w$$

$$Av = AD^2 A^T w$$

$$w = (AD^2 A^T)^{-1} Av$$

$$v_N = D^2 A^T (AD^2 A^T)^{-1} Av$$

Portanto,

$$Tv = v_T = v - \perp v = \left(I - D^2 A^T (AD^2 A^T)^{-1} A \right) v.$$

Desta forma, se definem as projeções tangencial e normal:

$$\begin{aligned} T(x) &= I - \perp(x) \\ \perp(x) &= D^2 A^T (AD^2 A^T)^{-1} A. \end{aligned} \tag{1.23}$$

Conexão Riemanniana Induzida

Seja $(N, \langle \cdot, \cdot \rangle')$ uma variedade Riemanniana com conexão Riemanniana ∇' dada pelo Teorema 1 (de Levi-Civita), e seja $(M, \langle \cdot, \cdot \rangle)$ uma subvariedade Riemanniana de N ($\langle \cdot, \cdot \rangle$ é a métrica induzida de N); então o teorema de Levi-Civita garante existência de uma conexão Riemanniana ∇ de M . Considerando $x \in M$, $X_x \in T_x M$ e $Y \in \Omega(M)$, estabelece-se a seguinte relação entre ∇ e ∇' :

Teorema 1.5

$$\nabla_X^Y = T \left(\nabla_{\bar{X}}^{\bar{Y}} \right) \tag{1.24}$$

onde \bar{X}, \bar{Y} são extensões locais a N dos campos X, Y .

Prova: Veja do Carmo [88], pág. 57 (exercício 3). ■

Corolário 1.1. Se c é uma curva em M e Y é um campo de vetores ao longo de c que é tangente à subvariedade M (ao longo de c), então

$$\frac{dY}{dt} = T \left(\frac{D'Y}{dt} \right).$$

Prova: Fazendo a extensão de Y a toda variedade N obtemos um campo \tilde{Y} em N tal que $\tilde{Y}(t) = Y(t)$ ao longo de c . Por propriedade de conexão, temos

$$\begin{aligned}\frac{D'\tilde{Y}}{dt} &= \nabla'_{\frac{dc}{dt}} \tilde{Y} = \nabla'_{\frac{dc}{dt}} Y := \frac{D'Y}{dt} \\ \frac{DY}{dt} &:= \nabla_{\frac{dc}{dt}} Y = T\left(\nabla'_{\frac{dc}{dt}} Y\right) = T\left(\frac{D'Y}{dt}\right),\end{aligned}$$

a segunda igualdade sendo decorrente do Teorema 5. ■

Corolário 1.2. Uma curva c em M é uma geodésica se e só se

$$\frac{D'}{dt} \left(\frac{dc}{dt} \right)$$

é sempre perpendicular a M . Em particular, se uma geodésica c em N está inteiramente contida em M , então c é uma geodésica em N .

Observação: Classicamente, a componente tangencial $T\left(\frac{D'}{dt} \left(\frac{dc}{dt}\right)\right) = T(c''(t))$ é chamada de vetor curvatura geodésica de c , enquanto

$$K_g(t) := \|T(c''(t))\|_{G(t)}$$

é chamada curvatura geodésica de c ; note que c é geodésica de M se e só se $K_g(t) = 0, \forall t$.

Exemplo 1.18. Retomemos o Exemplo 16. Uma curva

$$x: I \rightarrow M, \quad x(t) = (x_1(t), \dots, x_n(t))^T, \quad I = [a, b],$$

é geodésica de M se e só se

$$\frac{D'}{dt} \left(\frac{dx}{dt} \right) \in (T_{x(t)}M)^\perp, \quad \forall t \in I.$$

Calculamos, sucessivamente,

$$\begin{aligned} \frac{D'}{dt} \left(\frac{dx}{dt} \right) &= \sum_k \left\{ \frac{d}{dt} \left(\frac{dx_k}{dt} \right) + \sum_{i,j} \Gamma_{ij}^k \frac{dx_j}{dt} \frac{dx_i}{dt} \right\} \frac{\partial}{\partial x_k} = \\ &= \sum_k \left\{ \frac{d}{dt} \left(\frac{dx_k}{dt} \right) - \frac{1}{x_k} \left(\frac{dx_k}{dt} \right)^2 \right\} \frac{\partial}{\partial dx_k}. \end{aligned}$$

Usando o fato de que $(T_{x(t)}M)^\perp = \text{Im}(D^2(t)A^T)$ garante-se a existência de uma curva $\lambda: [0, w^*) \rightarrow \mathbf{R}^m$ tal que $\frac{D'}{dt} \left(\frac{dx}{dt} \right) = D^2 A^T \lambda(t)$, ou, em coordenadas,

$$\frac{\ddot{x}_k}{x_k} - \left(\frac{\dot{x}_k}{x_k} \right)^2 = x_k \alpha_k(t), \quad \text{onde } \alpha_k(t) = [A^T \lambda(t)]_k. \quad (1.25)$$

Observação: A geodésica poderia ter sido obtida pela clássica formulação do Cálculo Variacional, em que $x(t)$ é solução de:

$$\begin{aligned} \min_x \int_0^T \sqrt{\sum_i \frac{\dot{x}_i(t)^2}{x_i(t)^2}} dt, \quad \text{s.a.} \quad & Ax(t) = b \\ & x(0) = a, \dot{x}(0) = v, \end{aligned}$$

onde $x(\cdot)$ pertence a algum espaço adequado de funções suaves. O integrando é a norma de $\dot{x}(t)$ na métrica D^{-2} .

Observe que, conseqüência desta formulação, a curva $\lambda(t)$ acima definida é naturalmente produzida como multiplicador de Lagrange da restrição $Ax(t) = b$.

1.11. Gradiente e Hessiana

Gradiente

Denotemos por Ω o conjunto de campos de vetores em M , onde M é uma variedade Riemanniana.

Seja $f: M \rightarrow \mathbf{R}$ uma função diferenciável. Então o diferencial df da função f pertence ao dual Ω^* do espaço de campos de vetores Ω .

Definição 1.5. O gradiente, denotado por $\text{grad } f$, de uma função f é o campo de vetor metricamente equivalente ao diferencial df . Assim,

$$\langle \text{grad } f, X \rangle = df(X) = Xf, \quad (1.26)$$

$X \in \Omega$, onde Xf denota a derivada de f segundo a direção do campo X .

Se $x: U \subset \mathbf{R}^n \rightarrow M$ é um sistema de coordenadas locais em torno de $p \in M$, com

$$x(x_1, \dots, x_n) = q \in x(U) \quad \text{e} \quad \frac{\partial}{\partial x_i}(q) = dx_q(0, \dots, 1, 0, \dots, 0)$$

então

$$df = \sum_j \frac{\partial f}{\partial x_j} \frac{\partial}{\partial x_j} \quad \text{e} \quad \text{grad } f = \sum_i y_i \frac{\partial}{\partial x_i}$$

logo

$$y_i = \left\langle \text{grad } f, \frac{\partial}{\partial x_i} \right\rangle := df \left(\frac{\partial}{\partial x_i} \right) = \sum_j g_{ij} \frac{\partial f}{\partial x_j}$$

portanto

$$\text{grad } f(q) = G^{-1}(q)f'(q) \quad (1.27)$$

onde $G^{-1} = (g^{k\ell})$ e $f' = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right)^T$ é o vetor de derivadas parciais da função $f \circ x$.

Exemplo 1.19: $M = \mathbf{R}^n$ com métrica Euclidiana. Então $G^{-1} = I$ e, portanto, (27) coincide com o vetor das derivadas parciais.

Exemplo 1.20: $M = \text{int}(\mathbf{R}_+^n) = \{x \in \mathbf{R}^n; x > 0\}$ com a métrica dada por (7); assim,

$$\text{grad } f(x) = D^2 f(x) \quad (1.28)$$

onde $D = \text{diag}(x_1, \dots, x_n)$.

Exemplo 1.21: $M = \{x \in \mathbf{R}^n; e^T x = 1, x > 0\}$ com a métrica bi-invariante dada por (8); assim,

$$\text{grad } f(x) = (D - xx^T)Df'(x) \quad (1.29)$$

Exemplo 1.22: Seja A matriz $m \times n$, posto de A igual a $m < n$ e $S = \{x \in M; Ax = b, b \in \mathbf{R}^m\}$ onde M é dado pelo Exemplo 17 acima. Com a métrica induzida, temos que

$$\text{grad } f(x) = [I - D^2 A^T (AD^2 A^T)^{-1}] D^2 f'(x)$$

ou seja,

$$\text{grad } f(x) = D[I - DA^T (AD^2 A^T)^{-1} AD] Df'(x) \quad (1.30)$$

Definição 1.6. $p \in M$ é um ponto crítico de f se $\text{grad } f(p) = 0$.

Hessiana

Definição 1.7. Seja $f: M \rightarrow \mathbf{R}$ de classe C^r , $r \geq 2$. A hessiana de f , denotada por H^f , é definida como a derivada covariante do campo gradiente.

Para $p \in M$, $v \in T_p M$,

$$H_p^f: T_p M \rightarrow T_p M, \quad H_p^f \cdot v := \nabla_v^{\text{grad } f} f. \quad (1.31)$$

Lema 1.1. Para cada $p \in M$, o operador H_p^f é auto-adjunto, isto é,

$$\langle H_p^f \cdot v, w \rangle_p = \langle v, H_p^f \cdot w \rangle_p, \quad (1.32)$$

para todo $v, w \in T_p M$.

Prova: Decorrem da definição, página 17, os seguintes fatos:

- i) ∇ é compatível com a métrica $\Leftrightarrow X \langle Y, Z \rangle = \langle \nabla_X^Y + Z \rangle + \langle Y, \nabla_X^Z \rangle$
- ii) ∇ é simétrica $\Leftrightarrow \nabla_X^Y - \nabla_Y^X = [X, Y]$.

Fazendo $X = v$, $Y = w$ e $Z = \text{grad } f$ em (i), obtemos:

$$v \langle w, \text{grad } f \rangle = \langle \nabla_v^w, \text{grad } f \rangle + \langle w, H_p^f \cdot v \rangle$$

trocando $X = w$ e $Y = v$, obtemos

$$w \langle v, \text{grad } f \rangle = \langle \nabla_w^v, \text{grad } f \rangle + \langle v, H_p^f \cdot w \rangle$$

Agora,

$$\begin{aligned}
& \langle H_p^f \cdot v, w \rangle - \langle v, H_p^f \cdot w \rangle = (v \langle w, \text{grad } f \rangle - \langle \nabla_v^w, \text{grad } f \rangle) - \\
& - (w \langle v, \text{grad } f \rangle - \langle \nabla_w^v, \text{grad } f \rangle) = \\
& = [v(wf) - w(vf)] - [\langle \nabla_v^w - \nabla_w^v, \text{grad } f \rangle] = \\
& = (vw - wv)f - (\nabla_v^w - \nabla_w^v)f = \\
& = [v, w]f - [v, w]f = 0
\end{aligned}$$

onde na última igualdade usamos (ii). ■

Com o Lema 1, para cada $p \in M$, definimos uma forma quadrática simétrica

$$\begin{aligned}
H_p^f: T_p M \times T_p M &\rightarrow \mathbf{R} \\
H_p^f(v, w) &= \langle H_p^f \cdot v, w \rangle_p = \langle \nabla_v^{\text{grad } f}, w \rangle
\end{aligned}$$

ou mais geralmente

$$\begin{aligned}
H^f: \Omega \times \Omega &\rightarrow \mathbf{R} \\
H^f(X, Y) &= \langle \nabla_X^{\text{grad } f}, Y \rangle
\end{aligned} \tag{1.33}$$

Observação: A fórmula (33) tem a desvantagem de depender do conhecimento da métrica e da conexão, quando sabemos que uma determina a outra; assim, desenvolvemos o

Lema 1.2. Para todo $X, Y \in \Omega$

$$H^f(X, Y) = (XY - \nabla_X^Y)f \tag{1.34}$$

Prova:

$$X \langle Y, \text{grad } f \rangle = \langle \nabla_X^{\text{grad } f}, Y \rangle + \langle \nabla_X^Y, \text{grad } f \rangle$$

Logo,

$$H^f(X, Y) = X(Yf) - (\nabla_X^Y)f = (XY - \nabla_X^Y)f. \quad \blacksquare$$

Exemplo 1.23: $M = \mathbf{R}^n$ com a métrica usual. Então $\Gamma_{ij}^k = 0$ (veja pág. 32); logo,

$$\nabla_{\frac{\partial}{\partial x_j}}^{\frac{\partial}{\partial x_i}} = 0 \text{ e}$$

$$H^f \left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \right) = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

coincide com a hessiana usual.

Exemplo 1.24: $M = \text{int}(\mathbf{R}_+^n)$, com a métrica $g_{ij}(x) = \frac{\delta_{ij}}{x_i x_j}$. Então $\Gamma_{ij}^m = -\frac{1}{x_i} \delta_{im} \delta_{ij}$ (veja pág. 34). Assim,

$$\nabla_{\frac{\partial}{\partial x_i}}^{\frac{\partial}{\partial x_j}} = -\frac{\delta_{ij}}{x_i} \frac{\partial}{\partial x_i}.$$

Portanto,

$$H^f \left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \right) = \begin{cases} \frac{\partial^2 f}{\partial x_i \partial x_j} & \text{se } i \neq j \\ \frac{\partial^2 f}{\partial^2 x_i} + \frac{1}{x_i} \frac{\partial f}{\partial x_i} & i = j \end{cases} \quad (1.35)$$

Note que:

i) Se $f: M \rightarrow \mathbf{R}$, $f(x) = -\sum_{i=1}^n \log x_i$, então

$$\text{grad } f(x) = x \quad \text{e} \quad H_x^f = 0.$$

ii) Se $f(x) = \sum_i c_i x_i$ então

$$\text{grad } f(x) = D^2 c \quad \text{e} \quad H_x^f = \text{diag}(D^{-1} c)$$

onde $D = \text{diag}(x_1, \dots, x_n)$.

Exemplo 1.25: $M = \{x \in \mathbf{R}^n; e^T x = 1, x > 0\}$ com a métrica $g_{ij}(x) = \frac{1}{x_i x_j} [\delta_{ij} - \frac{1}{n}]$. Então

$$\Gamma_{ij}^m(x) = \delta_{ij} \left[\frac{\delta_{im}}{x_m} - \frac{x_m}{x_i} \right] \text{ (veja Exemplo 14). Assim,}$$

$$\nabla_{\frac{\partial}{\partial x_i}}^{\frac{\partial}{\partial x_j}} := \sum_{m=1}^n \Gamma_{ij}^m \frac{\partial}{\partial x_m} = \begin{cases} 0 & i \neq j \\ -\frac{1}{x_i} \frac{\partial}{\partial x_i} + \frac{1}{x_i} \sum_m x_m \frac{\partial}{\partial x_m} & i = j \end{cases}$$

Logo

$$H^f \left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \right) = \begin{cases} \frac{\partial^2 f}{\partial x_i \partial x_j} & i \neq j. \\ \frac{\partial^2 f}{\partial x_i^2} - \frac{1}{x_i} \sum_m x_m \frac{\partial f}{\partial x_m} + \frac{1}{x_i} \frac{\partial f}{\partial x_i}, & i = j \end{cases} \quad (1.36)$$

Observação: Se $p \in S$ é ponto crítico de f , nos exemplos acima, a hessiana H_p^f coincide com a hessiana de $f \circ x: U \subset \mathbf{R}^n \rightarrow \mathbf{R}$ onde $x: U \rightarrow S$ sistema de coordenadas locais em torno do ponto p .

A observação se generaliza no

Corolário 1.3. Se $p \in S$ é ponto crítico de f e $X, Y \in \Omega$ então

$$H_p^f(X_p, Y_p) = X_p(Y_p f).$$

Prova:

$$H_p^f(X_p, Y_p) = X_p(Y_p f) - \left\langle \nabla_{X_p}^Y, \text{grad } f(p) \right\rangle.$$

Como $\text{grad } f(p) = 0$, segue o corolário. ■

Funções Lineares e Quadráticas em Variedade Riemanniana

Em analogia com o \mathbf{R}^n , introduzimos a seguinte

Definição 1.8. Seja $f: M \rightarrow \mathbf{R}$ de classe C^2 . f é dita:

- (i) Linear se $H^f(p) = 0, \forall p \in M$;
- (ii) Quadrática se $H^f(p) = H$ constante $\neq 0, \forall p \in M$.

Exemplo 1.26. $M = \mathbf{R}^n$ com a métrica usual. Assim,

- (i) f é linear se e só se $\frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial x_j} \right) = 0, \forall i, j$. Logo

$$\frac{\partial f}{\partial x_j}(x) = a_j, \quad \forall j;$$

portanto

$$f(x) = \sum_{j=1}^n a_j x_j + b, \text{ com } a_j, b \in \mathbf{R}.$$

(ii) f é quadrática se e só se $\frac{\partial^2 f}{\partial x_i \partial x_j} = a_{ij}$, ou equivalentemente $\frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial x_j} \right) = a_{ij}$, que implica em $\frac{\partial f}{\partial x_j} = a_{ij} x_i + \varphi(x_j)$.

Como $\frac{\partial^2 f}{\partial x_j^2} = a_{jj}$, temos sucessivamente $\varphi'(x_j) = a_{jj}$, $\varphi(x_j) = a_{jj} x_j + b_j$, $\frac{\partial f}{\partial x_j} = a_{ij} x_i + a_{jj} x_j + b_j$, $f(x) = \sum_{i \neq j} a_{ij} x_i x_j + \sum_{j=1}^n a_{jj} x_j^2 + \sum_j b_j x_j + c$ e $f(x) = \frac{1}{2} x^T Q x + b^T x + c$, onde $Q = (a_{ij})$ é simétrica.

Exemplo 1.27. $M = \{x \in \mathbf{R}^n; x > 0\}$ com a métrica $g_{ij}(x) = \frac{\delta_{ij}}{x_i x_j}$; usando (35), temos

(i) f é linear se e só se $\frac{\partial^2 f}{\partial x_i \partial x_j} = 0$ e $\frac{\partial^2 f}{\partial x_j^2} + \frac{1}{x_j} \frac{\partial f}{\partial x_j} = 0$.

Como $\frac{\partial^2 f}{\partial x_i \partial x_j} = 0, \forall i \neq j$, tem-se $f(x) = \sum_{j=1}^n b_j f_j(x_j)$, onde cada f_j satisfaz

$$\frac{d^2 f_j}{dx_j^2} + \frac{1}{x_j} \frac{df_j}{dx_j} = 0$$

como $f_j(x_j) = \log(a_j x_j)$, $a_j > 0$, satisfaz esta última equação, segue-se portanto que f é linear se e só se

$$f(x) = \sum_{j=1}^n b_j \log(a_j x_j).$$

(ii) f é quadrática se e só se $\frac{\partial^2 f}{\partial x_i \partial x_j} = a_{ij}$ e $\frac{\partial^2 f}{\partial x_j^2} + \frac{1}{x_j} \frac{\partial f}{\partial x_j} = a_{jj}$. Consequentemente $\frac{\partial f}{\partial x_j} = a_{ij} x_i + \psi(x_j)$ o que implica em $\frac{\partial^2 f}{\partial x_j^2} + \frac{1}{x_j} \frac{\partial f}{\partial x_j} = \psi'(x_j) + a_{ij} \frac{x_i}{x_j} + \frac{\psi(x_j)}{x_j}$. Portanto $a_{ij} x_i = a_{jj} x_j - x_j \psi'(x_j) - \psi(x_j), \forall i \neq j$, o que é equivalente a $a_{ij} = 0$. Assim $f(x) = \sum_{j=1}^n f_j(x_j)$ onde cada f_j satisfaz:

$$\frac{d^2 f_j}{dx_j^2} + \frac{1}{x_j} \frac{df_j}{dx_j} = a_{jj}, \quad j = 1, \dots, n.$$

Para encontrar a solução geral desta equação, basta encontrar uma solução particular, pois para a equação homogênea associada, temos a solução do caso linear acima. Procuremos solução particular do tipo

$$f_j(x_j) = \beta x_j^\gamma$$

Por substituição obtemos $\gamma = 2$ e $\beta = \frac{a_{jj}}{4}$, ou seja $f_j(x_j) = \frac{a_{jj}}{4} x_j^2$.

Portanto f é quadrática se e só se

$$f(x) = \sum_{j=1}^n \frac{a_{jj}}{4} x_j^2 + \sum_{j=1}^n b_j \log(a_j x_j) + c$$

onde $c, a_{jj}, b_j, a_j \in \mathbf{R}$ e $a_j > 0$.

Hessiana em subvariedades

Como foi feito para o gradiente, consideremos a seguinte situação: Seja $(N, \langle \cdot, \cdot \rangle')$ variedade Riemanniana com conexão Riemanniana ∇' e seja $(M, \langle \cdot, \cdot \rangle)$ subvariedade Riemanniana, onde $\langle \cdot, \cdot \rangle$ é a métrica induzida de N ; já vimos que a conexão ∇ em M , dada pelo Teorema de Levi-Civita, satisfaz a relação

$$\nabla_X^Y = T_{(p)} \left(\nabla' \frac{\bar{Y}}{\bar{X}} \right) \quad (1.37)$$

para todo $p \in M, X, Y \in \Omega$, onde $T_{(p)}: T_p N \rightarrow T_p M$ é a projeção definida em (22). Se $\bar{f}: N \rightarrow \mathbf{R}$ e $f = \bar{f}|_M: M \rightarrow \mathbf{R}$, sabemos que

$$\text{grad } f(p) = T_{(p)}(\text{grad } \bar{f}(p)) \quad (1.38)$$

Interessa-nos determinar a relação entre a hessiana de \bar{f} e a hessiana de f .

De (37), temos

$$\begin{aligned} B: \Omega \times \Omega &\rightarrow (\Omega)^\perp \\ B(X, Y) &= \nabla' \frac{\bar{Y}}{\bar{X}} - \nabla_X^Y \end{aligned} \quad (1.39)$$

onde Ω^\perp denota os campos de vetores normais à subvariedade M .

Lema 1.3. A aplicação B definida por (39) é bilinear e simétrica.

Prova: Veja do Carmo [88], pág. 127. ■

Seja $p \in M$ e $\eta \in (T_p M)^\perp$. A aplicação $H_\eta: T_p M \times T_p M \rightarrow \mathbf{R}$ dada por

$$H_\eta(x, y) := \langle B(x, y), \eta \rangle$$

é, pelo Lema 3, uma forma bilinear simétrica.

Observe que à aplicação bilinear H_η fica associada uma aplicação linear auto-adjunta $S_\eta: T_p M \rightarrow T_p M$ por

$$\langle S_\eta(x), y \rangle = H_\eta(x, y) = \langle B(x, y), \eta \rangle \quad (1.40)$$

Teorema 1.6. Seja $p \in M$, $x \in T_p M$ e $\eta \in (T_p M)^\perp$. Seja $\tilde{\eta}$ uma extensão local de η normal a M . Então

$$S_\eta(x) = -T_{(p)}(\nabla'_x \tilde{\eta}). \quad (1.41)$$

Prova: Veja do Carmo [88], pág. 128. ■

De (38), podemos escrever

$$\text{grad } \bar{f} = \text{grad } f + \perp (\text{grad } \bar{f}).$$

Observe que $\perp (\text{grad } \bar{f}) \in (\Omega^\perp)$ e que uma extensão a N do campo $\text{grad } f$, denotada por $\overline{\text{grad } f}$, á dada por

$$\overline{\text{grad } f} = \text{grad } \bar{f} - \overline{(\perp (\text{grad } \bar{f}))} \quad (1.42)$$

onde $\overline{(\perp (\text{grad } \bar{f}))}$ é uma extensão do campo $\perp (\text{grad } \bar{f})$.

Teorema 1.7. Seja $p \in M$ e $x \in T_p M$. Então

$$H_p^f \cdot x = T_{(p)}(H_p^{\bar{f}} \cdot x) + S_{\perp(\text{grad } \bar{f})}(x) \quad (1.43)$$

onde S_n é dada por (41) com $\eta = \perp (\text{grad } \bar{f}(p))$.

Prova:

Usando a definição (7) e a extensão (42), temos:

$$\begin{aligned} H_p^f \cdot x &:= \nabla_x^{\text{grad } f} = T_{(p)} \left(\nabla_x^{\overline{\text{grad } f}} \right) = T_{(p)} \left(\nabla_x^{(\text{grad } \bar{f} - \perp(\overline{\text{grad } \bar{f}}))} \right) \\ &= T_{(p)} \left[\left(\nabla_x^{\text{grad } \bar{f}} \right) - T_{(p)} \left(\nabla_x^{\overline{\perp(\text{grad } \bar{f})}} \right) \right] \\ &= T_{(p)}(H_x^{\bar{f}}) - T_{(p)} \left(\nabla_x^{\overline{\perp(\text{grad } \bar{f})}} \right). \end{aligned}$$

Usando o Teorema 6 na última igualdade, concluímos a prova. ■

Para $M = F^{-1}(0)$, onde $F: N \rightarrow \mathbf{R}^m$ é de classe C^∞ e "0" é valor regular de F , então M é subvariedade Riemanniana de dimensão $(n - m)$. Neste caso, temos:

Corolário 1.4. Se $F = (F_1, \dots, F_m)$, $p \in M$ e $x \in T_p M$, então

$$S_{\perp(\text{grad } \bar{f})}(x) = - \sum_{j=1}^m \lambda_j(p) T_{(p)}(H_p^{F_j} \cdot x) \quad (1.44)$$

onde $\lambda(p)$ é dado por $\perp(\text{grad } \bar{f}(p)) = \sum_{j=1}^m \lambda_j(p) \text{grad } F_j(p)$.

Prova:

Note que $\{\text{grad } F_1(p), \dots, \text{grad } F_m(p)\}$ é base do $(T_p M)^\perp$; assim, existe $\lambda(p) \in \mathbf{R}^m$ (único) tal que

$$\perp(\text{grad } \bar{f}(p)) = \sum_{j=1}^m \lambda_j(p) \text{grad } F_j(p).$$

Para cada $q \in N$, o campo

$$\perp(\overline{\text{grad } \bar{f}})(q) := \sum_{j=1}^m \lambda_j(p) \text{grad } F_j(q)$$

é uma extensão do campo $\perp(\text{grad } \bar{f})$ normal a M .

Pelo Teorema 6, temos

$$\begin{aligned}
S_{\perp(\text{grad } \bar{f})}(x) &= -T_{(p)} \left(\nabla'_x \sum \lambda_j(p) \text{grad } F_j \right) \\
&= -T_{(p)} \left(\sum \left[\frac{\partial \lambda_j}{\partial x} \nabla'_x \text{grad } F_j + \lambda_j(p) \nabla'_x \text{grad } F_j \right] \right) \\
&= - \sum_j \lambda_j(p) T_{(p)} \left(\nabla'_x \text{grad } F_j \right),
\end{aligned}$$

pois $\frac{\partial \lambda_j}{\partial x} = 0$.

Usando a definição de hessiana, conclui-se o Corolário 4. ■

Corolário 1.5. Seja $M = F^{-1}(0)$, $p \in M$ e $x \in T_p M$. Então

$$H_p^f \cdot x = T_{(p)} \left[H_p^{\bar{f}} \cdot x - \sum_{j=1}^m \lambda_j(p) H_p^{F_j} \cdot x \right] \quad (1.45)$$

onde $\lambda_j(p)$ é dado em (44).

Prova:

Segue do Teorema 7 e do Corolário 4. ■

Exemplo 1.28: Seja $N = \text{int}(\mathbf{R}_+^n)$ com a métrica $g_{ij}(p) = \frac{\delta_{ij}}{p_i p_j}$, onde $p \in N$. Seja $F: N \rightarrow \mathbf{R}^m$, $F(p) = Ap - b$ com A matriz de posto $m < n$. Temos $F_j(p) = a_j^T p - b_j$ onde a_j é a j -ésima linha da matriz A . Suponha $M = F^{-1}(0)$. Se $\bar{f}(p) = \sum_{i=1}^n c_i p_i$ temos de (41)

$$H_p^{\bar{f}} = \text{diag}(D^{-1}c), \quad H_p^{F_j} = \text{diag}(D^{-1}a_j)$$

onde $D = \text{diag}(p_1, \dots, p_n)$.

Neste caso, $\lambda^T = c^T D^2 A^T (AD^2 A^T)^{-1}$ e $T_{(p)} = I - D^2 A^T (AD^2 A^T)^{-1} A$ e substituindo em (45) encontra-se a hessiana da função $f(y) = \sum_i c_i y_i$ restrita a $Ay = b$.

1.13. Variedade Completa e Teorema de Hopf/Rinow

Tendo em vista o objetivo de desenvolver algoritmos em variedades Riemannianas que façam uma minimização unidimensional ao longo de Geodésicas, é natural trabalhar com variedades em que as geodésicas estejam definidas para qualquer valor do parâmetro as variedades completas. Além disso, o que também as torna úteis é o fato (Teorema de Hopf/Rinow) de que dados dois pontos quaisquer existe sempre uma geodésica que os liga que minimiza o comprimento entre todas as curvas diferenciáveis por partes que os ligam.

Definição 1.9. Uma variedade Riemanniana M é (geodesicamente) completa se para todo $p \in M$, as geodésicas $\gamma(t)$ que partem de p estão definidas para todos os valores do parâmetro $t \in \mathbf{R}$.

Exemplo 1.29: $M = \mathbf{R}^n$ com a métrica euclidiana $g_{ij}(x) = \delta_{ij}$ é completa, pois dados

$$\gamma(0) = x \in M \quad \text{e} \quad \dot{\gamma}(0) = v \in T_x M$$

vimos que a i -ésima componente da geodésica, que é

$$\gamma_i(t) = v_i t + x_i \quad i = 1, \dots, n$$

está definida para todo valor do parâmetro $t \in \mathbf{R}$.

Exemplo 1.30: $M = \text{int}(\mathbf{R}_+^n)$ com a métrica $g_{ij}(x) = \frac{\delta_{ij}}{x_i x_j}$ é completa, pois dados

$$\gamma(0) = x \in M \quad \text{e} \quad \dot{\gamma}(0) = v \in T_x M$$

vimos em (17) que a i -ésima componente da geodésica é

$$\gamma_i(t) = x_i \exp\left(\frac{v_i}{x_i} t\right), \quad i = 1, \dots, n$$

que está definida para todo parâmetro $t \in \mathbf{R}$.

Exemplo 1.31: $M = \{x \in \mathbf{R}^n; e^T x = 1, x > 0\}$ com a métrica $g_{ij}(x) = \frac{1}{x_i x_j} (\delta_{ij} - \frac{1}{n})$ é completa, pois dados

$$\gamma(0) = x \in M \quad \text{e} \quad \dot{\gamma}(0) = v \in T_x M$$

vimos em (21) que a i -ésima componente da geodésica é

$$\gamma_i(t) = x_i \exp \left[\left(\frac{v_i}{x_i} - \frac{v_1}{x_1} \right) t \right] / \left(\sum_{k=1}^n x_k \exp \left[\left(\frac{v_k}{x_k} - \frac{v_1}{x_1} \right) t \right] \right)$$

que está definida para todo $t \in \mathbf{R}$.

No que se segue, consideraremos M conexa.

É conveniente introduzir uma distância numa variedade Riemanniana (não necessariamente completa) M como segue. São dados $p, q \in M$ e seja $\gamma: [a, b] \rightarrow M$ diferenciável por partes tal que $\gamma(a) = p$ e $\gamma(b) = q$.

Definição 1.10. A distância $d(p, q)$ é definida por

$$d(p, q) := \inf_{\gamma} \int_a^b \|\dot{\gamma}(t)\| dt. \quad (1.46)$$

Teorema 1.8. Com a distância d em (46), M é espaço métrico, isto é:

- 1) $d(p, r) \leq d(p, q) + d(r, q)$
- 2) $d(p, q) = d(q, p)$
- 3) $d(p, q) \geq 0$ e $d(p, q) = 0 \Leftrightarrow p = q$.

Prova: Veja do Carmo [88], pág. 146. ■

Teorema 1.9. A topologia induzida por d em M coincide com a topologia inicial de M .

Prova: Veja do Carmo [88], pág. 146. ■

Teorema 1.10 (Hopf/Rinow). Seja M uma variedade Riemanniana e seja $p \in M$. As seguintes afirmações são equivalentes:

- 1) Os limitados e fechados de M são compactos.
- 2) M é completa como espaço métrico.
- 3) M é geodesicamente completa.
- 4) Para todo $q \in M$ existe uma geodésica γ ligando p a q com $d(p, q) =$ comprimento de γ , isto é, o ínfimo em (46) é atingido por uma geodésica.

Prova: Veja do Carmo [88] pág. 147. ■

Exemplo 1.32: $M = \text{int}(\mathbf{R}_+^n)$ com a métrica $g_{ij}(x) = \frac{\delta_{ij}}{x_i x_j}$. Dados $p, q \in M$ existe uma única geodésica ligando p a q , pois se $\gamma_i(t) = p_i \exp\left(\frac{v_i t}{p_i}\right)$ e $\alpha_i(t) = p_i \exp\left(\frac{w_i t}{p_i}\right)$ são as i -ésimas componentes das geodésicas que satisfazem $\gamma_i(0) = \alpha_i(0) = p_i$ e $\gamma_i(1) = \alpha_i(\bar{t}) = q_i$, então

$$q_i = p_i \exp\left(\frac{v_i}{p_i}\right) = p_i \exp\left(\frac{w_i \bar{t}}{p_i}\right) \Rightarrow v_i = \bar{t} w_i.$$

Vamos obter a expressão da distância nesta variedade:

i) $q_i = p_i \exp\left(\frac{v_i}{p_i}\right) \Rightarrow v_i = \log(q_i/p_i) p_i$

ii) $\frac{\dot{\gamma}_i(t)}{\gamma_i(t)} = \frac{v_i}{p_i} \Rightarrow \left(\frac{\dot{\gamma}_i(t)}{\gamma_i(t)}\right)^2 = \frac{v_i^2}{p_i^2} = [\log(q_i/p_i)]^2$

Assim,

$$\begin{aligned} d(p, q) &= \int_0^1 \sqrt{\sum_{i,j} g_{ij}(\gamma(t)) \dot{\gamma}_i(t) \dot{\gamma}_j(t)} dt = \int_0^1 \sqrt{\sum_{i,j} \frac{\delta_{ij}}{\gamma_i(t) \gamma_j(t)} \dot{\gamma}_i(t) \dot{\gamma}_j(t)} dt \\ &= \int_0^1 \sqrt{\sum_{i=1}^n \left(\frac{\dot{\gamma}_i(t)}{\gamma_i(t)}\right)^2} dt. \end{aligned}$$

Logo

$$d^2(p, q) = \sum_{i=1}^n [\log(q_i/p_i)]^2 \tag{1.47}$$

Exemplo 1.33: $M = \{x \in \mathbf{R}^n; e^T x = 1, x > 0\}$ com a métrica $g_{ij}(x) = \frac{1}{x_i x_j} [\delta_{ij} - \frac{1}{n}]$. Dados $p, q \in M$ mostra-se que

$$d^2(p, q) = \sum_{i=1}^n \left[\log(p_i/q_i) - \frac{1}{n} \sum_{j=1}^n \log(p_j/q_j) \right]^2 \quad (1.48)$$

veja Karmarkar [90], pág. 59.

Corolário 1.6. Se M é compacta, então M é completa.

Corolário 1.7. Uma subvariedade fechada de uma variedade Riemanniana completa é completa na métrica induzida; em particular, as subvariedades fechadas de um espaço euclidiano são completas.

Observação: Do Corolário 7, segue que o problema

$$\begin{aligned} \min f(x) \\ \text{s.a. } F(x) = 0 \end{aligned} \quad (1.49)$$

pode ser visto como um problema particular de minimizar uma função definida em uma variedade Riemanniana completa, desde que $F^{-1}(0)$ seja conexa.

O teorema a seguir fornece condição suficiente para a conexidade de $F^{-1}(0)$.

Teorema 1.11. Para $m \leq n$, seja $F: \mathbf{R}^n \rightarrow \mathbf{R}^m$ de classe C^r , $r \geq 2$ tal que

$$\sup\{\|(DF(x)DF(x)^T)^{-1}\|; x \in \mathbf{R}^n\} < \infty.$$

Então $F^{-1}(0)$ é uma subvariedade conexa do \mathbf{R}^n .

Prova: Veja Diener [87], pág. 182. ■

1.14. Teorema de Toponogov e a lei dos cossenos

É razoável esperar que a taxa de convergência de um método ao longo de geodésicas dependa da velocidade de afastamento das geodésicas que partem de cada ponto $p \in M$. Pode-se mostrar que, localmente, tal velocidade de afastamento depende do sinal da curvatura seccional de M no ponto p ; isto é, o seguinte esquema se verifica:

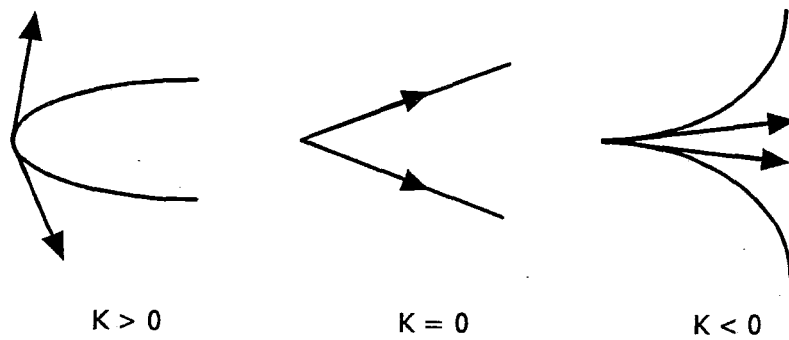


Figura 9

Numa variedade Riemanniana denominamos de triângulo geodésico o conjunto de três geodésicas $\gamma_1, \gamma_2, \gamma_3, \gamma_i: [0, l_i] \rightarrow M$ tal que

$$\gamma_i(l_i) = \gamma_{i+1}(0) \text{ (módulo 3) e } l_i + l_{i+1} \geq l_{i+2}.$$

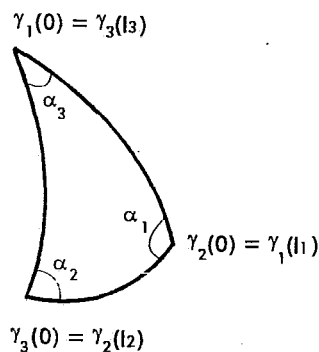


Figura 10

Denotaremos por $\alpha_i, 0 \leq \alpha_i \leq \pi$ o ângulo entre $-\dot{\gamma}_i(l_i)$ e $\dot{\gamma}_{i+1}(0)$.

Quando $M = \mathbb{R}^2$ com uma métrica $g_{ij}(x) = \text{cte}$ que não depende do ponto $x \in M$, então M tem curvatura seccional – chamada curvatura Gaussiana-nula; as geodésicas são

retas. Neste caso, dados três pontos distintos a, b, c em M , tais pontos determinam um único triângulo geodésico T de M com vértices a, b, c . Seja α o ângulo do vértice c e sejam A, B, C os comprimentos dos lados opostos aos vértices a, b, c , respectivamente. Então vale a lei dos cossenos:

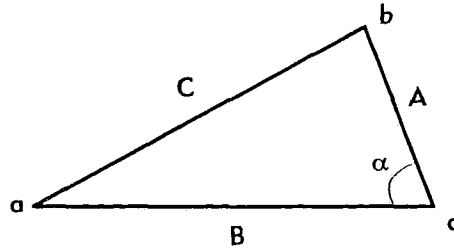


Figura 11

$$A^2 + B^2 - 2AB \cos \alpha = C^2 \quad (1.50)$$

Procuramos estabelecer uma relação do tipo (50) em um triângulo geodésico em função do sinal da curvatura seccional de M .

Teorema 1.12. Seja M uma variedade Riemanniana completa, simplesmente conexa com curvatura seccional $k \leq 0$. Sejam a, b, c três pontos de M . Tais pontos determinam um único triângulo geodésico T de M com vértices a, b, c . Seja α o ângulo do vértice c e sejam A, B, C os comprimentos dos lados opostos aos vértices a, b, c , respectivamente. Então vale

$$A^2 + B^2 - 2AB \cos \alpha \leq C^2 \quad (< C^2 \text{ se } k < 0) \quad (1.51)$$

Prova: Veja do Carmo [88], pág. 259. ■

Para obter uma relação do tipo lei dos cossenos para curvatura seccional não-negativo ($k \geq 0$), precisamos do Teorema de Toponogov, que está enunciado em do Carmo [88], pág. 235.

Teorema 1.13 (Toponogov). Seja M uma variedade Riemanniana completa com curvatura seccional $k \geq H$. Sejam γ_1 e γ_2 segmentos de geodésicas normalizadas em M com $\gamma_1(0) = \gamma_2(0)$. Indiquemos por $M^2(H)$ uma variedade de dimensão dois com curvatura constante H .

Admitamos que a geodésica γ_1 é minimizante e que, se $H > 0$, o comprimento de γ_2 é menor ou igual a π/\sqrt{H} . Consideremos em $M^2(H)$ duas geodésicas normalizadas $\bar{\gamma}_1, \bar{\gamma}_2$, tais que $\bar{\gamma}_1(0) = \bar{\gamma}_2(0)$, $L(\gamma_i) = L(\bar{\gamma}_i) = \ell_i$, $i = 1, 2$, e $\text{ang}(\dot{\bar{\gamma}}_1(0), \dot{\bar{\gamma}}_2(0)) = \text{ang}(\dot{\gamma}_1(0), \dot{\gamma}_2(0))$. Então

$$d(\gamma_1(\ell_1), \gamma_2(\ell_2)) \leq d(\bar{\gamma}_1(\ell_1), \bar{\gamma}_2(\ell_2)). \quad (1.52)$$

Observação:

i) No Teorema 13 a notação $L(\gamma_i) = L(\bar{\gamma}_i) = \ell_i$ é o comprimento da geodésica γ_i .

ii) Para o próximo teorema, adotaremos a seguinte notação

$\gamma_1, \gamma_2, \gamma_3$ triângulo geodésico, com cada γ_i normalizada tais que γ_1, γ_2 são minimizantes; além disso, denotemos por $\alpha = \text{ang}(\dot{\gamma}_1(0), -\dot{\gamma}_3(\ell_3))$, $A = L(\gamma_1)$, $B = L(\gamma_3)$ e $C = L(\gamma_2)$.

Teorema 1.14 (Lei dos cossenos). Seja M uma variedade Riemanniana completa com curvatura seccional $k \geq 0$. Seja $\{\gamma_1, \gamma_2, \gamma_3\}$ triângulo geodésico normalizado tal que γ_1, γ_2 segmentos de geodésicas minimizantes. Com a notação acima, vale

$$C^2 \leq A^2 + B^2 - 2AB \cos \alpha \quad (1.53)$$

Prova: Seja M^2 a variedade afim, gerada pelos vetores $\dot{\gamma}_1(0)$, $\dot{\gamma}_3(\ell_3)$ e transladado para $\gamma_1(0)$, munido da métrica constante e igual a de M no ponto $\gamma_1(0)$; assim, M^2 é variedade Riemanniana de dimensão dois e curvatura seccional nula.

Note que as geodésicas de M^2 são retas; em particular considere

$$\tilde{\gamma}_1(t) = t\dot{\gamma}_1(0),$$

$$\tilde{\gamma}_3(t) = -t\dot{\gamma}_3(\ell_3).$$

Em M^2 temos o triângulo $\{\gamma_1(0), \tilde{\gamma}_1(A), \tilde{\gamma}_3(B)\}$, que satisfaz:

$$A = d(\gamma_1(0), \tilde{\gamma}_1(A)), \quad B = d(\gamma_1(0), \tilde{\gamma}_3(B)) \quad \text{e} \quad \alpha = \text{ang}(\dot{\tilde{\gamma}}_1(0), \dot{\tilde{\gamma}}_3(B)).$$

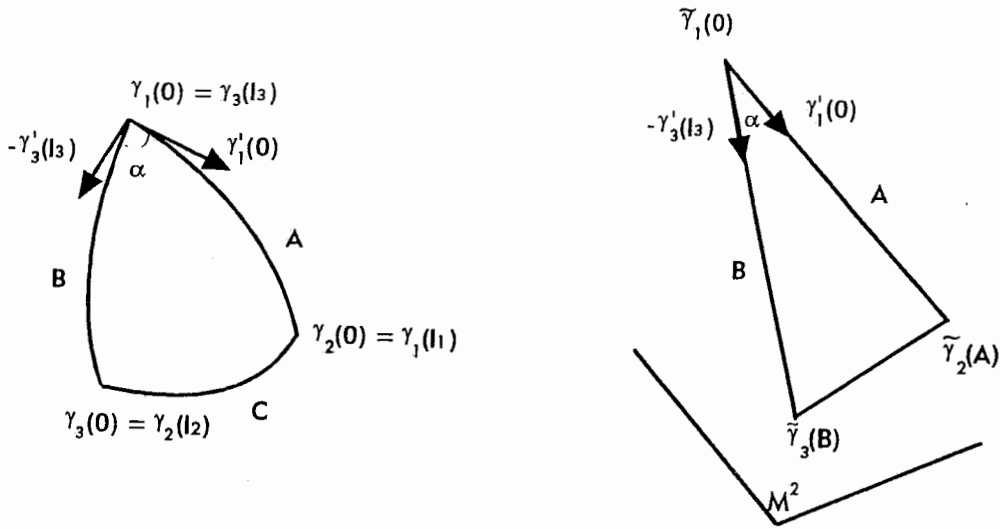


Figura 12

Usando a lei dos cossenos em M^2 , obtemos:

$$d^2(\tilde{\gamma}_1(A), \tilde{\gamma}_3(B)) = A^2 + B^2 - 2AB \cos \alpha.$$

Pelo Teorema de Toponogov, temos

$$C^2 = d^2(\gamma_1(A), \gamma_3(0)) \leq d^2(\tilde{\gamma}_1(A), \tilde{\gamma}_3(B)) = A^2 + B^2 - 2AB \cos \alpha$$

que conclui a prova. ■

CAPÍTULO 2

FUNDAMENTOS: CONDIÇÕES DE OTIMALIDADE E ELEMENTOS DE ANÁLISE CONVEXA

2.1. Introdução

O elo entre a Geometria e a Programação Não Linear se estabeleceu primeiramente através das condições de otimalidade e dos conceitos básicos de análise convexa. Este capítulo trata destes pontos, utilizando parte do material desenvolvido no capítulo anterior. Em geral os resultados se assemelham formalmente aos clássicos, válidos em variedades lineares.

2.2. Condições de Otimalidade

Seja M uma variedade Riemanniana completa e $f: M \rightarrow \mathbf{R}$ de classe C^r , $r \geq 1$. Nesta seção derivaremos condições de otimalidade para o problema

$$\min_{x \in M} f(x). \quad (2.1)$$

Como consequência, deduziremos as condições de otimalidade para o problema clássico

$$\begin{aligned} &\min f(x) \\ &\text{s.a} \\ &h(x) = 0 \\ &x \in \mathbf{R}^n, \end{aligned}$$

onde $h: \mathbf{R}^n \rightarrow \mathbf{R}^m$ é classe C^r e “0” valor regular para função h .

Teorema 2.1. (Condição necessária de primeira ordem).

Seja $f: M \rightarrow \mathbf{R}$ de classe C^1 . Se x^* é ponto de mínimo, então $\text{grad } f(x^*) = 0$.

Prova: Basta mostrar que $\langle \text{grad } f(x^*), v \rangle = 0$ para todo $v \in T_{x^*} M$.

Dado $v \in T_{x^*} M$. Seja $\gamma: \mathbf{R} \rightarrow M$ geodésica tal que $\gamma(0) = x^*$ e $\dot{\gamma}(0) = v$. Então $t = 0$ é ponto de mínimo local para a função

$$h(t) = f(\gamma(t));$$

portanto $h'(0) = 0$; desde que $h'(0) = \langle \text{grad } f(x^*), v \rangle$, segue o resultado. ■

Como corolário do Teorema 1, obtemos o teorema dos Multiplicadores de Lagrange.

Corolário 2.1. (Multiplicadores de Lagrange).

Seja $h: \mathbf{R}^n \rightarrow \mathbf{R}^m$ de classe C^r com "0" valor regular de h . Se x^* é uma solução local para

$$\min f(x)$$

s.a

$$h(x) = 0$$

$$x \in \mathbf{R}^n,$$

então existe $\lambda^* \in \mathbf{R}^m$ tal que

$$f'(x^*) - \sum_{i=1}^m \lambda_i^* h'_i(x^*) = 0,$$

onde $h = (h_1, \dots, h_m)^T$.

Prova: Seja $M = h^{-1}(0)$, então M , com a métrica induzida do \mathbf{R}^n , é variedade Riemanniana completa e, além disso (Capítulo 1),

a) $\text{grad } f(x) = T(x)f'(x)$

b) $(T_x M)^\perp = \left\{ \sum_{i=1}^m v_i h'_i(x); v_i \in \mathbf{R}, i = 1, \dots, m \right\}$.

Assim, se x^* é ponto de mínimo local, segue do Teorema 1 que $\text{grad } f(x^*) = 0$.

(a) Implica que $f'(x^*) \in (T_{x^*}M)^\perp$ e de (b) existe $\lambda^* \in \mathbf{R}^m$ tal que

$$f'(x^*) = \sum_{i=1}^m \lambda_i^* h'_i(x^*),$$

o que mostra o corolário. ■

Teorema 2.2. (Condição necessária de segunda ordem).

Seja $f: M \rightarrow \mathbf{R}$ de classe C^2 . Se x^* é ponto de mínimo local, então

$$v^T H_{x^*}^f v \geq 0, \quad \forall v \in T_{x^*}M. \quad (2.2)$$

Prova: Dado $v \in T_{x^*}M$. Seja $\gamma: \mathbf{R} \rightarrow M$, $\gamma(0) = x^*$, $\dot{\gamma}(0) = v$ geodésica e defina $h: \mathbf{R} \rightarrow \mathbf{R}$, $h(t) = f(\gamma(t))$. Então

$$\text{i) } h'(t) = \text{grad } f(\gamma(t))^T \dot{\gamma}(t) \Rightarrow h'(0) = 0$$

$$\text{ii) } h''(t) = \dot{\gamma}(t)^T H_{\gamma(t)}^f \dot{\gamma}(t) + \text{grad } f(\gamma(t))^T \ddot{\gamma}(t) = \dot{\gamma}(t)^T H_{\gamma(t)}^f \dot{\gamma}(t) \text{ pois } \dot{\gamma} \text{ é campo paralelo; logo, } h''(0) = v^T H_{x^*}^f v.$$

Desde que x^* é ponto de mínimo local, segue que $t = 0$ é ponto de mínimo local para h ; portanto, $h''(0) \geq 0$ o que mostra o teorema. ■

Corolário 2.2. Sejam h e x^* como no Corolário 2. Então existe $\lambda^* \in \mathbf{R}^m$ tal que

$$\text{i) } f'(x^*) - \sum_i \lambda_i^* h'_i(x^*) = 0.$$

$$\text{ii) Se } L(x^*) := H_{x^*}^f - \sum_i \lambda_i^* H_{x^*}^{h_i}, \text{ então } v^T L(x^*) v \geq 0 \text{ para todo } v \in \mathbf{R}^n \text{ tal que } Dh(x^*) \cdot v = 0.$$

Prova: Seja M como na prova do Corolário 2. (i) é este Corolário. Para mostrar (ii), basta observar que:

$$\text{(a) } T_{x^*}M = \{v \in \mathbf{R}^n; Dh(x^*)v = 0\}.$$

(b) $\bar{f} = f|_M$, então o Corolário 1.4 implica,

$$\begin{aligned} v^T H_{x^*}^{\bar{f}} \cdot v &= v^T T(x^*) [H_{x^*}^f - \sum \lambda_i^* H_{x^*}^{h_i}] v \\ &= v^t [H_{x^*}^f - \sum_i \lambda_i^* H_{x^*}^{h_i}] v \\ &= v^T L(x^*) v, \text{ para } v \text{ tal que } Dh(x^*)v = 0. \end{aligned}$$

O corolário segue aplicando o Teorema 2 à função $\bar{f}: M \rightarrow \mathbf{R}$. ■

Teorema 2.3 (Condição suficiente de 2ª ordem).

Seja $f: M \rightarrow \mathbf{R}$ de classe C^2 . Se x^* é ponto de M satisfazendo:

- i) $\text{grad } f(x^*) = 0$.
- ii) $H_{x^*}^f$ definida positivo

então x^* é ponto de mínimo local de f .

Prova: Dado $v \in T_{x^*}M$, $\|v\| = 1$.

Seja η o menor auto-valor da matriz $H_{x^*}^f$, que é positivo. Por Taylor, temos

$$f(\gamma(t)) = f(x^*) + \frac{t^2}{2} v^T H_{x^*}^f v + 0(t^2),$$

onde $\gamma: (-\varepsilon, \varepsilon) \rightarrow M$ é uma curva satisfazendo $\gamma(0) = x^*$ e $\dot{\gamma}(0) = v$. Assim,

$$f(\gamma(t)) \geq f(x^*) + \frac{\eta}{2} t^2 + 0(t^2)$$

e, portanto, existe t_0 que não depende de v , tal que para $0 \leq t \leq t_0$ tem-se que

$$f(\gamma(t)) \geq f(x^*). \quad \blacksquare$$

Corolário 2.3. Seja h como nos corolários anteriores e suponha que exista $x^* \in \mathbf{R}^n$, $\lambda^* \in \mathbf{R}^m$ tais que $h(x^*) = 0$ e $f'(x^*) - \sum_i \lambda_i^* h'_i(x^*) = 0$. Suponha também que se $y \in \mathbf{R}^n$, $Dh(x^*)y = 0$

então $y^T L(x^*)y > 0$, onde $L(x^*) = H_{x^*}^f - \sum \lambda_i^* H_{x^*}^{h_i}$. Então x^* é ponto de mínimo local de f sujeito a $h(x) = 0$.

Prova: Análogo aos casos anteriores. ■

2.3. Convexidade em Variedades Riemannianas

2.3.1. Introdução

Esta seção foi motivada pelo trabalho de T. Rapcsak [91], o qual generaliza o conceito de função convexa em variedades diferenciáveis munidas da métrica induzida do \mathbf{R}^n ; basicamente, obtém condições de 1ª e 2ª ordens de convexidade.

Estabelecemos os resultados obtidos por Rapcsak, para funções definidas em uma variedade Riemanniana completa, que contém, em particular, as consideradas por aquele autor. Além disso, mostramos que, no caso da variedade ser compacta, todas as funções convexas são constantes.

2.3.2. Convexidade de Conjuntos e de Funções

Existem diversas generalizações na Geometria Riemanniana que estendem o conceito de convexidade em \mathbf{R}^n ; discutiremos alguns de importância para a programação matemática.

Definição 2.1: Dizemos que $A \subset M$ é totalmente convexo se para quaisquer dois pontos p e q de A , não necessariamente distintos, as geodésicas ligando p e q estão inteiramente contidas em A .

Exemplo 2.1. $M = \mathbf{R}^n$, com a métrica usual, qualquer conjunto convexo, no sentido clássico, é totalmente convexo.

Exemplo 2.2. Se $p \in M$ e existe um laço geodésico não trivial em p , isto é, uma geodésica $\gamma: [a, b] \rightarrow M$ tal que $\gamma(a) = p = \gamma(b)$ e $\gamma(t) \neq p$ para algum $t \in [0, 1]$, então o conjunto

$A = \{p\}$ não é totalmente convexo (logo, em geral, conjuntos unitários não são totalmente convexos).

Definição 2.2: Dizemos que $A \subset M$ é convexo se para todo par de pontos p e q em A existe uma geodésica minimal ligando p e q , contida em A .

Exemplo 2.3. M e conjuntos unitários são conjuntos convexos.

Exemplo 2.4. $M = S^n$, com métrica induzida do \mathbf{R}^{n+1} , e A o hemisfério inferior aberto é convexo, enquanto um hemisfério fechado não é convexo.

Definição 2.3: $f: M \rightarrow \mathbf{R}$ é dita função convexa se sua restrição a qualquer geodésica de M é uma função convexa da reta, isto é, se $\gamma: \mathbf{R} \rightarrow M$ é geodésica, então

$$f \circ \gamma: \mathbf{R} \rightarrow \mathbf{R} \text{ é convexa.}$$

Teorema 2.4: $f: M \rightarrow \mathbf{R}$ é convexa se e somente se para todo $\gamma: [a, b] \rightarrow M$ segmento de geodésica e para qualquer $\lambda \in [0, 1]$, verifica-se

$$f(\gamma((1 - \lambda)a + \lambda b)) \leq (1 - \lambda)f(\gamma(a)) + \lambda f(\gamma(b)) \quad (2.3)$$

Prova: Segue direto da definição. ■

Note que a definição acima é a generalização natural da definição clássica de função convexa para $M = \mathbf{R}^n$ munido da métrica usual, pois, dados p e q , a geodésica ligando p e q é $\gamma: [0, 1] \rightarrow M$,

$$\begin{aligned} \gamma(\lambda) &= p + \lambda(q - p) = (1 - \lambda)p + \lambda q \\ f(\gamma(\lambda)) &= f((1 - \lambda)p + \lambda q) \leq (1 - \lambda)f(p) + \lambda f(q). \end{aligned}$$

Para $a \in \mathbf{R}$, definimos o conjunto de nível M^a , $M^a = \{x \in M; f(x) \leq a\}$.

Corolário 2.4: Se $f: M \rightarrow \mathbf{R}$ é convexa, então M^a é totalmente convexo.

Prova: Dados $p, q \in M^a$. Seja $\gamma: [0, b] \rightarrow M$ geodésica tal que $\gamma(0) = p$ e $\gamma(b) = q$. Mostremos que $\gamma(t) \in M^a$, para todo $t \in [0, b]$.

Existe $\lambda \in [0, b]$ tal que $t = \lambda b$. Pelo Teorema 4, segue

$$\begin{aligned} f(\gamma(t)) &= f(\gamma((1 - \lambda)0 + \lambda b)) \leq (1 - \lambda)f(\gamma(0)) + \lambda f(\gamma(b)) \\ &= (1 - \lambda)f(p) + \lambda f(q) \leq (1 - \lambda)a + \lambda a = a. \end{aligned}$$

Logo, $\gamma(t) \in M^a$. ■

Teorema 2.5: Se $f: M \rightarrow \mathbf{R}$ é convexa, então f é localmente Lipschitziana, portanto, contínua.

Prova: A prova é idêntica ao caso clássico e será omitida. ■

No que segue, $f: M \rightarrow \mathbf{R}$ denotará função de classe C^r , $r \geq 1$; em geral, estaremos interessados no caso $r \geq 2$.

2.3.3. Condições Necessárias e Suficientes para Convexidade de Funções

Teorema 2.6 (Primeira Ordem): $f: M \rightarrow \mathbf{R}$ é convexa se e só se, para qualquer ponto $p \in M$, $\gamma: \mathbf{R} \rightarrow M$ geodésica com $\gamma(0) = p$, vale a desigualdade

$$f(\gamma(t)) - f(p) \geq t \langle \text{grad } f(p), \dot{\gamma}(0) \rangle \quad (2.4)$$

Prova: (\Rightarrow) Definimos $h: \mathbf{R} \rightarrow \mathbf{R}$, $h(t) = f(\gamma(t))$. De ser f convexa e classe C^r segue que h é convexa e de classe C^r , portanto

$$h(t) - h(0) \geq th'(0).$$

Agora observemos que $h'(0) = \langle \text{grad } f(p), \dot{\gamma}(0) \rangle$ o que mostra (4).

(\Leftarrow) De modo análogo, usando (4) mostra-se que h é convexa. ■

Corolário 2.5: Se $f: M \rightarrow \mathbf{R}$ é convexa, então todo ponto crítico de f é ponto de mínimo global.

Prova: Seja $p \in M$ ponto crítico de f , isto é,

$$\text{grad } f(p) = 0 \tag{2.5}$$

dado $q \in M$, $q \neq p$, pelo Teorema de Hopf-Rinow, existe uma geodésica $\gamma: \mathbf{R} \rightarrow M$ e $b \in \mathbf{R}_+$ tais que

$$\gamma(0) = p \quad \text{e} \quad \gamma(b) = q.$$

Usando (4), temos que

$$f(\gamma(b)) - f(\gamma(0)) \geq b \langle \text{grad } f(p), \dot{\gamma}(0) \rangle$$

de (5), segue

$$f(q) \geq f(p), \text{ para todo } q \in M,$$

o que mostra que p é ponto de mínimo global de f . ■

O corolário a seguir fornece condição suficiente para que uma função convexa seja constante.

Corolário 2.6: Seja M compacta e $f: M \rightarrow \mathbf{R}$ convexa. Então f é constante.

Prova: Da compacidade de M e da continuidade de f , segue que f tem ponto de máximo e mínimo, os quais são pontos críticos de f ; da convexidade de f e do Corolário 5, segue que o ponto de máximo é ponto de mínimo e, portanto, f é constante. ■

Observamos que Bishop-O'Neill [69] usa o conceito de funções convexas para obter informações sobre a topologia da variedade; neste trabalho, através de técnicas de sistemas dinâmicos, estabelecem o resultado abaixo, do qual o Corolário 6 é um caso particular.

“Se M é variedade Riemanniana completa tendo volume finito, então todas as funções convexas em M são constantes.”

Teorema 2.7 (Segunda Ordem): Seja $f: M \rightarrow \mathbf{R}$ de classe C^r , $r \geq 2$. f é convexa se e somente se, para todo $p \in M$ a hessiana de f em p ,

$$H_p^f: T_p M \times T_p M \rightarrow \mathbf{R},$$

é semi-definida positiva.

Prova: (\Rightarrow) Dados $p \in M$ e $v \in T_p M$, queremos mostrar que

$$H_p^f(v, v) \geq 0.$$

Seja $\gamma: \mathbf{R} \rightarrow M$ a geodésica (única) tal que

$$\gamma(0) = p \quad \text{e} \quad \dot{\gamma}(0) = v.$$

Definindo $h: \mathbf{R} \rightarrow \mathbf{R}$, $h(t) = f(\gamma(t))$ sabemos que h é convexa e de classe C^r ; da análise convexa clássica, sabemos que $h''(t) \geq 0$.

Agora

$$\begin{aligned} h'(t) &= \langle \text{grad } f(\gamma(t)), \dot{\gamma}(t) \rangle \\ 0 \leq h''(t) &= \left\langle \nabla_{\dot{\gamma}(t)}^{\text{grad}} f(\gamma(t)), \dot{\gamma}(t) \right\rangle + \left\langle \text{grad } f(\gamma(t)), \nabla_{\dot{\gamma}(t)} \dot{\gamma}(t) \right\rangle \\ &= \left\langle \nabla_{\dot{\gamma}(t)}^{\text{grad}} f(\gamma(t)), \dot{\gamma}(t) \right\rangle = H_{\gamma(t)}^f(\dot{\gamma}(t), \dot{\gamma}(t)) \end{aligned} \quad (2.6)$$

devido a que $\nabla_{\dot{\gamma}(t)} \dot{\gamma}(t) = 0$ (γ é geodésica) e, na última igualdade, devido ao Lema 1.1 (item 1.11). Em particular, para $t = 0$ temos

$$H_p^f(v, v) \geq 0$$

como queríamos mostrar.

(\Leftarrow) Dado $p \in M$, seja $\gamma: \mathbf{R} \rightarrow M$ uma geodésica tal que $\gamma(0) = p$; basta mostrar que $h: \mathbf{R} \rightarrow \mathbf{R}$, $h(t) = f(\gamma(t))$ é convexa; da análise clássica, significa mostrar que $h''(t) \geq 0$. De (6) e da hipótese da semi-definição positiva da hessiana, segue que $h''(t) \geq 0$, o que mostra que f é convexa. ■

Corolário 2.7 (Teorema 5.1 de Rapcsak [91]): Suponha que

$$M = \{x; h_j(x) = 0, j = 1, \dots, n - k, x \in \mathbf{R}^n\}$$

seja conexa; seja $f: M \rightarrow \mathbf{R}$ de classe C^2 e suponha que $\{\text{grad } h_j(x); j = 1, \dots, n-k, x \in M\}$ seja linearmente independente. Então f é convexa se e somente se a matriz

$$H_x^g f(x, \mu(x)) = \left(H_x f + \sum_{j=1}^{n-k} \mu_j(x) H_x h_j \right) \Big|_{T_x M}, \quad (2.7)$$

é semi-definida positiva para todo $x \in M$, onde H_x^f e $H_x h_j$ são as hessianas euclidianas das funções f e h_j , $j = 1, \dots, n-k$, e $\mu(x)$ é a componente ortogonal do $\text{grad } f(x)$.

Prova: Com a hipótese dos gradientes linearmente independentes, M é variedade Riemanniana completa, com a métrica induzida do \mathbf{R}^n ; do Corolário 1.5, segue que $H_x^g f(x, \mu(x))$ em (7) é a hessiana de $f|_M$; pelo Teorema 7 segue o corolário. ■

CAPÍTULO 3

MÉTODOS GEODÉSICOS

3.1. Introdução

A associação entre métricas e direções de descida é uma ferramenta utilizada desde os primórdios da programação matemática contínua. Parcialmente explorada nos textos, por exemplo Luenberger [84] e Poliak [87], esta idéia é sistematicamente explorada neste capítulo; além disso, discutiremos a convergência de métodos geodésicos que usam a direção do gradiente, associado à métrica, para busca unidimensional ao longo da curva geodésica.

3.2. Uma Visão de Métodos Primais Através de Métricas Riemannianas

Partindo da propriedade de que o gradiente depende da métrica (item 1.11) e a solução do subproblema que define direções de descida (veja abaixo) são equivalentes, revisamos o resultado de que a maioria dos métodos primais clássicos são métodos de gradiente em alguma métrica (Riemanniana) adequada; além disso, verifica-se que as direções utilizadas pelos algoritmos afim-escala de Dikin, projetivo de Karmarkar e o multiplicativo de Eggermont podem ser enquadrados como gradiente. Através desta associação é possível gerar métodos (Direções) a partir de qualquer métrica Riemanniana.

Direção de Máxima Descida

Consideremos o problema geral da programação não-linear

$$\min_{x \in M \subset \mathbf{R}^n} f(x) \tag{3.1}$$

onde $f: \mathbf{R}^n \rightarrow \mathbf{R}$ é uma função de classe C^r , $r \geq 1$, e M é uma variedade Riemanniana.

Nesta estrutura, define-se classicamente o problema gerador de direção de descida, relativo a (1):

$$\begin{aligned} \min \quad & df(x; v) \\ \text{s.a.} \quad & \langle v, v \rangle_x = 1 \\ & v \in T_x M \end{aligned} \tag{3.2}$$

onde $df(x; v)$ é a derivada direcional de f em x na direção $v \in T_x M$. Observamos que a solução (única) de (2) depende do produto interno definido no espaço tangente de M no ponto x , isto é, depende da métrica Riemanniana definida em M .

Desde que o gradiente é metricamente equivalente ao diferencial de f , então

$$df(x; v) = \langle \text{grad } f(x), v \rangle$$

e pela desigualdade de Cauchy-Schwartz, segue que $V^*(x) = -\frac{\text{grad } f(x)}{\|\text{grad } f(x)\|}$ é a solução de (2).

Nas próximas seções apresentamos, em 2.1, alguns métodos primais clássicos, sob a perspectiva de métodos de gradiente, seguindo-se em 2.2, algoritmos de pontos interiores. Limitamo-nos sempre às restrições lineares.

3.2.1. Métodos Clássicos

Progração não-linear irrestrita

Consideremos o caso particular de (1) dado por

$$\min_{x \in \mathbb{R}^n} f(x) \tag{3.3}$$

temos o resultado trivial:

Teorema 3.1:

1. Suponha $f \in C^1$. Então a direção de Cauchy é a direção oposta do gradiente onde o \mathbb{R}^n é munido da métrica euclidiana.

2. Suponha $f \in C^3$ e admita que $M := \{x \in \mathbf{R}^n; f''(x) \text{ é definida positiva}\}$ é não vazio. Então o método de Newton aplicado ao problema (1) é um método de gradiente, onde M é munido da métrica definida pela Hessiana $f''(x)$.

Prova: Mostremos (2). Como M é aberto, segue que $T_x M = \mathbf{R}^n$ para todo $x \in M$.

Para $x \in M$,

$$\langle \cdot, \cdot \rangle_x : T_x M \times T_x M \rightarrow \mathbf{R}$$

$$\langle v, w \rangle_x = \langle f''(x)v, w \rangle,$$

defina uma métrica Riemanniana em M , pois $f \in C^3$; além disso, $v \in T_x M$,

$$df(x, v) = \langle f'(x), v \rangle = \langle f''(x)[f''(x)]^{-1} f'(x), v \rangle$$

$$= \langle [f''(x)]^{-1} f'(x), v \rangle_x.$$

Da unicidade do campo gradiente, segue que a direção de Newton é a direção oposta à do gradiente. ■

Observação: De modo análogo, o teorema pode ser estendido a uma matriz $G(x)$ simétrica definida positiva; por exemplo, as utilizadas nos métodos de quase-Newton.

Programação não-linear com restrições de igualdade lineares

Trata-se de métodos primais que resolvem a classe de problemas dada por

$$\begin{aligned} \min f(x) \\ \text{s.a. } Ax = b \\ x \in \mathbf{R}^n \end{aligned}$$

onde A é uma matriz $m \times n$, de posto m ($m < n$) e $b \in \mathbf{R}^m$.

Munindo o \mathbf{R}^n com uma métrica Riemanniana qualquer, para o sistema de coordenadas canônicas, a cada $x \in \mathbf{R}^n$ está associada uma matriz simétrica definida positiva $n \times n$ $G(x)$;

o problema gerador de direção de descida em x :

$$\begin{aligned} \min \quad & \langle f'(x), v \rangle \\ \text{s.a.} \quad & \langle G(x)v, v \rangle = 1 \\ & Av = 0, v \in \mathbf{R}^n \end{aligned} \tag{3.4}$$

As condições de otimalidade de (4) fornecem a solução (a menos de constante)

$$V^*(x) = -G^{-1/2}[I - G^{-1/2} A^T(AG^{-1} A^T)^{-1} AG^{-1/2}]G^{-1/2} f'(x) \tag{3.5}$$

Teorema 3.2. A direção (5) é oposta ao gradiente da restrição da função f à subvariedade das restrições munida da métrica induzida do espaço ambiente.

Prova: Seja $\tilde{f} = f|_{Ax=b}$. Com a métrica induzida, temos

$$\text{grad } \bar{f}(x) = T(x) \text{grad } f(x),$$

onde $T(x)$ é a projeção sobre o espaço tangente da subvariedade no ponto x . Esta é dada por

$$T(x) = I - A^*(x)[AA^*(x)]^{-1} A,$$

onde $A^*(x)$ é a adjunta da matriz A ; isto é,

$$A^*(x) = G^{-1}(x)A^T.$$

Portanto,

$$T(x) = I - G^{-1} A^T(AG^{-1}(x)A^T)^{-1} A.$$

Usando o fato que $\text{grad } f(x) = G^{-1}(x)f'(x)$, obtemos $-\text{grad } \bar{f}(x) = V^*(x)$. ■

3.2.2. Métodos Primais de Pontos Interiores

Estamos interessados na seguinte classe de problemas:

$$\begin{aligned} \min f(x) \\ \text{s.a } Ax = b \\ x \geq 0 \end{aligned} \tag{3.6}$$

onde $f: \mathbf{R}^n \rightarrow \mathbf{R}$ é uma função de classe C^r , ($r \geq 1$). A é uma matriz $m \times n$ de posto m ($m < n$) e $b \in \mathbf{R}^m$.

Métodos primais de pontos interiores é a classe de métodos que resolvem o problema (6) gerando uma seqüência $\{x^k\}$ satisfazendo:

- i) $x^{k+1} = x^k + t_k V^k$, onde $V^k \in A^\perp$ (espaço nulo de A) e $t_k \in \mathbf{R}_+$ de modo que $f(x^{k+1}) < f(x^k)$;
- ii) $Ax^k = b$
- iii) $x_j^k > 0$, para $j = 1, \dots, n$.

Os algoritmos de Dikin (1967) e Karmarkar (1984) são os principais representantes desta classe de métodos. A seguir mostraremos que o de Dikin é método gradiente, isto é, $V^k = -\text{grad } f(x^k)$ em uma métrica conveniente; além disso, obteremos a direção de Karmarkar como uma representação da direção de Dikin no espaço tangente do simplex unitário.

Algoritmo de Dikin

O algoritmo associa a cada $x \in \mathbf{R}^n$, $Ax = b$, $x > 0$, a direção de busca no espaço nulo de A ,

$$V(x) = -D[I - DA^T(AD^2A^T)^{-1}AD]Df'(x) \tag{3.7}$$

onde $D = D_x := \text{diag}(x_1, \dots, x_n)$.

A interpretação da direção afim-escala (7) como o (simétrico do) gradiente correspondente a uma determinada métrica Riemanniana é um resultado obtido por Karmarkar e apresentado por Bayer e Lagarias (1989).

Teorema 3.3. A direção afim-escala (7) é oposto ao gradiente da restrição da função f à subvariedade das restrições do problema (6) munida da métrica induzida do \mathbf{R}_+^n , dada no Exemplo 1.10 (item 1.5).

Prova: Veja Exemplo 1.22 (item 1.11). ■

A direção de Karmarkar derivada de Dikin

Karmarkar [84] mostra que o problema (6) é equivalente a

$$\begin{aligned} \min f(x) \\ \text{s.a. } Ax = 0 \\ e^T x = 1 \\ x \geq 0 \end{aligned} \tag{3.8}$$

onde f, A são como em (6) e, além disso, $e = (1, \dots, 1)^T \in \mathbf{R}^n$ satisfaz $Ae = 0$.

O algoritmo associa a cada $x \in \mathbf{R}^n$, $Ax = 0$, $e^T x = 1$ e $x > 0$, a direção de busca

$$V_k(x) = -[D - xx^T][I - DA^T(AD^2AT)^{-1}AD]Df'(x) \tag{3.9}$$

que pertence ao espaço tangente da variedade das restrições no ponto x .

O que faremos a seguir é utilizar a base do espaço tangente do simplex unitário no ponto x , obtida no Exemplo 1.11 (item 1.5), para obter uma representação adequada $Z(x)$ do espaço tangente, isto é, tal que

$$T_x S = \{Z(x)y; y \in \mathbf{R}^n\}.$$

Do Exemplo 1.11, temos que $V = \{v_1, \dots, v_{n-1}\}$, onde $v_i = e_i - x$ é base do $T_x S$; assim, para trabalharmos com vetores de n coordenadas, consideremos a matriz $Z(x) = (v_1, \dots, v_n)$ cuja i -ésima coluna é o elemento v_i da base V e $v_n = e_n - x$. Reescrevemos $Z(x)$ como

$$Z(x) = I - xe^T \tag{3.10}$$

Teorema 3.4. A direção (9) do método projetivo de Karmarkar é dada pela $Z(x)$ -representação da direção (7) do método afim-escala de Dikin.

Prova: Dado $x \in \text{int}(\mathbf{R}_+^n)$ tal que $Ax = 0$ e $e^T x = 1$, seja $V(x)$ direção de Dikin, dada por (7). Então:

$$\begin{aligned} Z(x)V(x) &= -(I - xe^T)D[I - DA^T(AD^2A^T)^{-1}AD]Df'(x) \\ &= -[D - xx^T][I - DA^T(AD^2A^T)^{-1}AD]Df'(x) \\ &= V_k(x), \end{aligned}$$

onde usamos a igualdade $e^T D = x^T$. ■

Métodos multiplicativos

São métodos de pontos interiores propostos por Eggermont [90] para resolver o problema

$$\begin{aligned} \min f(x) \\ \text{s.a } x \geq 0 \end{aligned} \tag{3.11}$$

Dentre as famílias de direções propostas, Iusem [95] analisa a convergência do método descrito por

$$x^{k+1} = x^k - \lambda_k d^k,$$

onde

$$d_j^k := x_j^k (f'(x^k))_j, \quad j = 1, \dots, n \tag{3.12}$$

Propomo-nos mostrar que a direção (12) é a direção do gradiente relativo a uma determinada métrica Riemanniana definida em $\text{int}(\mathbf{R}_+^n)$.

Seja

$$h: \text{int}(\mathbf{R}_+^n) \rightarrow \mathbf{R}, \quad h(x) = \sum_{i=1}^N x_i \log x_i.$$

Considere int (\mathbf{R}_+^n) munido da métrica definido pela hessiana da função h ; assim, para cada $x \in \text{int}(\mathbf{R}_+^n)$, a matriz dos coeficientes da métrica é dada por

$$G(x) = \text{diag} \left(\frac{1}{x_1}, \dots, \frac{1}{x_n} \right). \quad (3.13)$$

Teorema 3.5. A direção (12) é o gradiente da função f relativa à métrica (13) definida em int (\mathbf{R}_+^n) .

Prova: Usando (1.26) (item 1.11), obtemos

$$(\text{grad } f(x))_j = x_j (f'(x))_j. \quad \blacksquare$$

Demonstraremos a seguir que a importante propriedade de variedade ser completa não é verificada no caso aqui tratado.

Teorema 3.6. int (\mathbf{R}_+^n) munido da métrica (13) é uma variedade Riemanniana incompleta.

Prova: $M := \text{int}(\mathbf{R}_+^n)$ com a métrica usual induzida do \mathbf{R}^n é uma variedade Riemanniana incompleta, pois suas geodésicas são semi-retas que não estão definidas para todo valor do parâmetro.

Basta então mostrar que

$$g = (g_1, \dots, g_n): M \rightarrow \text{int}(\mathbf{R}_+^n)$$

$$g_i(x) = \frac{1}{4} x_i^2$$

é uma isometria; desde que g é um difeomorfismo, é equivalente mostrar que o jacobiano

$$Jg(x): T_x M \rightarrow T_x(\text{int}(\mathbf{R}_+^n))$$

é uma isometria linear. Dados

$$v, w \in T_x M,$$

tem-se

$$\begin{aligned} \langle Jg(x).v, Jg(x)w \rangle_{g(x)} &= \sum_i \frac{4}{x_i^2} \left(\frac{1}{2} x_i v_i \right) \left(\frac{1}{2} x_i w_i \right) \\ &= \sum_i v_i w_i = \langle v, w \rangle. \quad \blacksquare \end{aligned}$$

Dada a matriz $A \in \mathbf{R}^{m \times n}$ de posto completo $m < n$ e para $b \in \mathbf{R}^m$, consideremos a subvariedade $S = \{x \in \text{int}(\mathbf{R}_+^n); Ax = b\}$ munida da métrica induzida.

De modo análogo ao Exemplo 1.17 (item 1.10), temos definido as projeções tangencial e normal dadas, respectivamente, por

$$\begin{aligned} \top(x) &= I - \perp(x) \\ \perp(x) &= DA^T(ADA^T)^{-1}A. \end{aligned}$$

Teorema 3.7. Seja $f: S \rightarrow \mathbf{R}$ diferenciável. O campo gradiente de f é dado por

$$\text{grad } f(x) = [D - DA^T(ADA^T)^{-1}AD]f'(x). \quad (3.14)$$

Prova: Segue da definição de gradiente e da expressão para $\perp(x)$. \blacksquare

Corolário 3.1. A direção de Karmarkar, no simplex unitário munido da métrica induzida (13), é o oposto do campo gradiente da função quadrática $f(x) = \frac{1}{2} \sum_i c_i x_i^2$.

Prova: Para $A = e^T$, temos que $x \in S$, $ADA^T = e^T x = 1$ e de (14) segue

$$\text{grad } f(x) = [D - xx^T]Dc = -V_k(x). \quad \blacksquare$$

3.3. Métodos de Descida Geodésica

A partir da conexão estabelecida entre métricas e direções de gradiente, a construção de um método primal de descida precisará da definição de uma curva sobre a qual se fará a

busca de mínimo local. Observe que a trajetória “natural” que é intrinsecamente ligado à estrutura Riemanniana, é a geodésica. Assim o método de descida, proposto por Luenberger [72] com fins de obter a taxa de convergência do algoritmo de gradiente projetado, é a opção coerente para as direções consideradas.

Gabay [82] estendeu o trabalho de Luenberger para o problema

$$\begin{aligned} \min f(x) \\ x \in M \end{aligned} \tag{3.15}$$

onde $f: M \rightarrow \mathbf{R}$ é de classe C^1 e $M \subset \mathbf{R}^n$ é uma variedade diferenciável com métrica induzida do \mathbf{R}^n .

Considere então o Algoritmo A a seguir:

Algoritmo A (Gradiente com busca exata)

1. Dado x_k , calcule o grad $f(x_k)$ no plano tangente à superfície das restrições M ;
2. Determine a geodésica (única) $x(t)$, $t \geq 0$, de M que verifique $x(0) = x_k$ e $\dot{x}_k(0) = -\text{grad } f(x_k)$;
3. Minimize $f(x(t))$, $t \geq 0$, obtendo t_k e faça $x_{k+1} = x(t_k)$.

Com a hipótese de conjunto de nível $M^0 = \{x \in M; f(x) \leq f(x_0)\}$ limitado, Gabay provou o

Teorema 3.8. A seqüência $\{x_k\}$ construída pelo algoritmo A (descida geodésica) está bem definida; ou é sempre finita, terminando em um ponto crítico, ou é infinita e qualquer ponto de acumulação é um ponto crítico. Se os valores críticos de f são distintos, então a seqüência $\{x_k\}$ convêrge a um ponto crítico.

Consideraremos também a convergência do algoritmo do gradiente nas versões de busca aproximada de Armijo e passo em um intervalo fixo.

Algoritmo B (Regra de Armijo).

O passo 3 é substituído por

$$t_k := 2^{-i_k} \bar{t},$$

onde $\bar{t} > 0$ é dado e i_k é o menor inteiro positivo tal que

$$f(x(t_k)) \leq f(x_k) - \alpha t_k \|\text{grad } f(x_k)\|^2,$$

sendo $\alpha \in (0, \frac{1}{2})$.

Gabay [82] observa que sua prova é válida para regra de Armijo.

Algoritmo C.

O passo 3 é substituído por:

são dados $\delta_1 > 0$ e $\delta_2 > 0$ tais que

$$\delta_1 \Gamma + \delta_2 < 1,$$

onde Γ é a constante de Lipschitz associado ao campo gradiente de f . Escolha

$$t_k \in \left(\delta_1, \frac{2}{\Gamma}(1 - \delta_2) \right). \quad (3.16)$$

Burachik [95] estuda a convergência do Algoritmo C para o caso em que $M = \mathbf{R}^n$ e f convexa.

Teorema 3.9. Seja $f: M \rightarrow \mathbf{R}$ de classe C^1 , M completa e gradiente Γ -Lipschitziano. Seja $\{x_k\}$ a seqüência gerada pelo Algoritmo C. Então:

(i) Existe uma constante $\beta > 0$ tal que

$$f(x_{k+1}) \leq f(x_k) - \beta t_k^2 \|\text{grad } f(x_k)\|^2$$

(ii) se o conjunto de nível M^0 é limitado, então a seqüência é fracamente convergente ao conjunto de pontos críticos, isto é, se verifica

(a) a seqüência $\{x_k\}$ é limitada;

(b) $\lim_{k \rightarrow \infty} d(x_{k+1}, x_k) = 0$;

(c) qualquer ponto de acumulação da seqüência $\{x_k\}$ é ponto crítico.

Na seção 3.5 provaremos a convergência da seqüência, retirando a hipótese de M^0 limitado, mas acrescentando a “convexidade” de f e a curvatura seccional não-negativa de M .

Para ficar mais claro o Algoritmo C, recordemos que se $\gamma: \mathbf{R} \rightarrow M$ é uma geodésica, e denotando por $P_\gamma(t): T\gamma(0)M \rightarrow T_{\gamma(t)}M$ o transporte paralelo ao longo da geodésica γ , que identifica o espaço tangente de M em $\gamma(0)$ com o espaço tangente em $\gamma(t)$, então $P_\gamma(t)$ é uma isometria linear para cada t fixo.

Denotemos por $L(t) := \int_0^t \|\dot{\gamma}(s)\| ds$ o comprimento da curva γ entre o ponto $\gamma(0)$ e $\gamma(t)$; assim, $\gamma: \mathbf{R} \rightarrow M$ é geodésica se e somente se $L(t) = \beta t$ para algum $\beta > 0$.

Definição 3.1. $f: M \rightarrow \mathbf{R}$ é dita de gradiente Lipschitziana, com constante Γ , se para todos $p, q \in M$ e $\gamma: [0, a] \rightarrow M$ geodésica com $\gamma(0) = p$ e $\gamma(a) = q$, se verifica

$$\|\text{grad } f(\gamma(t)) - P_{\gamma(t)} \text{grad } f(p)\| \leq \Gamma L(t) \quad (3.17)$$

para todo $t \in [0, a]$.

Lema 3.1. Sejam $p, q \in M$ e $\gamma: [0, a] \rightarrow M$ geodésica ligando p a q . Se $f: M \rightarrow \mathbf{R}$ é de classe C^1 , então

$$f(q) = f(p) + a \langle \text{grad } f(p), \dot{\gamma}(0) \rangle + \int_0^a \langle \text{grad } f(\gamma(t)) - P_{\gamma(t)} \text{grad } f(p), P_{\gamma(t)} \dot{\gamma}(0) \rangle dt \quad (3.18)$$

Prova: Do Teorema Fundamental do Cálculo, temos

$$f(q) - f(p) = \int_0^a \frac{d}{dt} f(\gamma(t)) = \int_0^a \langle \text{grad } f(\gamma(t)), \dot{\gamma}(t) \rangle dt.$$

Do fato de γ' ser campo paralelo e $P_{\gamma(t)}$ isometria, segue

$$\dot{\gamma}(t) = P_{\gamma(t)} \dot{\gamma}(0) \text{ e } \langle \text{grad } f(p), \dot{\gamma}(0) \rangle = \langle P_{\gamma(t)} \text{grad } f(p), P_{\gamma(t)} \dot{\gamma}(0) \rangle$$

e substituindo na igualdade acima, obtém-se (18). ■

Prova do Teorema: Mostremos (i)

De $\dot{\gamma}(0) = -\text{grad } f(x_k)$, Lema 1 segue

$$f(x_{k+1}) = f(x_k) - t_k \|\text{grad } f(x_k)\|^2 + \int_0^{t_k} \langle \text{grad } f(\gamma(t)) - P_{\gamma(t)} \text{grad } f(x_k), -P_{\gamma(t)} \text{grad } f(x_k) \rangle dt.$$

Usando a desigualdade de Cauchy-Schwartz juntamente com a hipótese de gradiente Γ -Lipschitziano e o fato de $P_{\gamma(t)}$ ser uma isometria linear, segue

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) - t_k \|\text{grad } f(x_k)\|^2 + \Gamma \frac{t_k^2}{2} \|\text{grad } f(x_k)\|^2 \\ &\leq f(x_k) - \left(\frac{1}{t_k} - \frac{\Gamma}{2} \right) \|\text{grad } f(x_k)\|^2 t_k^2. \end{aligned}$$

Com t_k escolhido como em (16), temos

$$0 < \frac{\delta_2 \Gamma}{2} \left(\frac{1}{1 - \delta_2} \right) = \beta \leq \frac{1}{t_k} - \frac{\Gamma}{2}$$

o que mostra (i).

Mostremos (ii):

A seqüência $\{f(x_k)\}$ é decrescente e da compacidade de M^0 , segue que existe

$$\lim_{k \rightarrow \infty} f(x_k) = \ell^* \leq f(x_n).$$

Assim,

$$\begin{aligned} \sum_{k=0}^n t_k^2 \|\text{grad } f(x_k)\|^2 &\leq \frac{1}{\beta} \sum_{k=0}^n (f(x_k) - f(x_{k+1})) = \frac{1}{\beta} (f(x_0) - f(x_{n+1})) \\ &\leq \frac{1}{\beta} (f(x_0) - \ell^*). \end{aligned}$$

Portanto, $\lim_{k \rightarrow \infty} t_k \|\text{grad } f(x_k)\| = 0$; como $\delta_1 < t_k$, segue que

$$\lim_{k \rightarrow \infty} \|\text{grad } f(x_k)\| = 0 \tag{3.19}$$

Temos então que:

- (a) segue da hipótese de que M^0 é limitado e de (i),
- (b) $d(x_{k+1}, x_k) \leq L(\gamma_k) = t_k \|\text{grad } f(x_k)\|$

Logo

$$\lim_{k \rightarrow \infty} d(x_{k+1}, x_k) \leq \lim_{k \rightarrow \infty} t_k \|\text{grad } f(x_k)\| = 0.$$

- (c) Seja x^* ponto de acumulação da seqüência $\{x_k\}$; assim, existe uma subsequência $\{x_{k_j}\}$ tal que

$$\lim_{j \rightarrow \infty} x_{k_j} = x^*.$$

Da continuidade do gradiente e de (19) segue que x^* é ponto crítico. ■

3.4. Exemplo

Algoritmo de descida geodésica no simplex unitário

Estamos interessados em resolver o problema

$$\begin{aligned} \min \quad & c^T x \\ \text{s.a} \quad & e^T x = 1 \\ & x \geq 0 \end{aligned} \tag{3.20}$$

Aplicaremos o Algoritmo A (busca exata) e mostraremos propriedades teóricas. Embora de solução trivial, o problema (20) tem servido a diversos autores como paradigma para a análise de complexidade de métodos de pontos interiores, no que se refere ao número de iterações. Citemos principalmente, Meggido-Shub [89], Karmarkar [90], Anstreicher [91] e Ye [94].

Obteremos, inicialmente, uma proposição semelhante à B.2 de Megiddo-Shub [89]. Em seguida, mostraremos que a função custo, ao longo da geodésica que passa pelo centro do simplex com velocidade dada pela direção de Karmarkar (ou seja, oposta ao gradiente), é

estritamente decrescente; além disso, fornecemos uma estimativa do tamanho do passo t_ε que garante, em uma iteração do Algoritmo A, uma ε -solução.

No que se segue, o vetor custo será da forma $c = (0, \dots, 0, c_{k+1}, \dots, c_n)^T$ onde $c_{k+1}, \dots, c_n > 0$, chamado por Karmarkar de normalizado; além disso, o simplex $\bar{S} = \{x \in \mathbf{R}_+^n; e^T x = 1\}$ é o fecho da variedade $S = \{x \in \text{int}(\mathbf{R}_+^n); e^T x = 1\}$ com a métrica bi-invariante.

Teorema 3.10. Dado $x \in S$. Seja $\gamma: \mathbf{R} \rightarrow S$ geodésica tal que $\gamma(0) = x$ e $\dot{\gamma}(0) = -\text{grad } f(x)$.

Então:

(a) Para $i, j \leq k$ e $t \in \mathbf{R}_+$,

$$\frac{\gamma_i(t)}{\gamma_j(t)} = \frac{x_i}{x_j}.$$

(b) $\lim_{t \rightarrow +\infty} \gamma(t) = \frac{1}{\sum_{i=1}^k x_i} (x_1, \dots, x_k, 0, \dots, 0)^T$,

que é uma solução do problema (20).

Em particular, se $x = \frac{1}{n} e$ (centro do simplex), então a geodésica tende ao centro da face ótima do problema.

Prova:

(a) Segue direto da expressão para a geodésica (1.20), (item 1.8).

(b) Decorre do fato de que $\lim_{t \rightarrow +\infty} \gamma_1(t) = \frac{x_1}{\sum_{j=1}^k x_j}$ e que $\gamma_i(t) = \frac{x_i}{x_1} \gamma_1(t) \exp(-c_i x_i t)$. ■

Teorema 3.11. Seja $x = \frac{1}{n} e$ (centro do simplex) e $\gamma: \mathbf{R} \rightarrow S$ geodésica como no Teorema 10. Então a função custo, ao longo da geodésica γ , $h: \mathbf{R}_+ \rightarrow \mathbf{R}$, $h(t) = \sum_{i=1}^n c_i \gamma_i(t)$ é monótona decrescente.

Prova: Se $x = \frac{1}{n} e$, então $\gamma_i(t) = \frac{1}{a(t)} \exp(-\frac{c_i}{n} t)$, onde

$$a(t) := \sum_{j=1}^n \exp\left(-\frac{c_j}{n} t\right).$$

Assim,

$$h(t) = \frac{1}{a(t)} \sum_{i=1}^n c_i \exp\left(-\frac{c_i}{n} t\right) = -n \frac{a'(t)}{a(t)}.$$

Logo,

$$h'(t) = \frac{n}{[a(t)]^2} [(a'(t))^2 - a''(t)a(t)].$$

Agora,

$$\begin{aligned} (a'(t))^2 - a''(t)a(t) &= \\ &= -\frac{1}{n^2} \left[\sum_{i,j=1}^n (c_i - c_j)^2 \exp[-(c_i + c_j)\frac{t}{n}] + \sum_{i=1}^n c_i^2 \exp\left(-\frac{c_i}{n} t\right) \right] < 0. \end{aligned}$$

Portanto, $h'(t) < 0$ que demonstra o teorema. ■

Corolário 3.2. Dado $\varepsilon > 0$. O Algoritmo A, aplicado ao problema (20), encontra uma solução aproximada verificando

$$\frac{c^T x_k}{c^T \frac{e}{n}} < \varepsilon \quad (3.21)$$

em uma iteração; uma estimativa do tamanho do passo pode ser dada por

$$t_\varepsilon = \frac{n}{\min\{c_i; c_i > 0\}} \log\left(\frac{n}{\varepsilon}\right).$$

Prova: Os Teoremas 10 e 11 garantem encontrar uma ε -solução para o problema (20) com uma iteração do Algoritmo A; basta então analisar o quociente em (21).

$$\begin{aligned} \frac{c^T \gamma(t)}{c^T \frac{e}{n}} &= \frac{n}{a(t)} \frac{1}{c^T e} \sum_{i=1}^n c_i \exp\left(-\frac{c_i}{n} t\right) \\ &\leq \frac{1}{c^T e} \left(n \sum_{i=1}^n c_i \right) \exp(-\min\{c_i; c_i > 0\} \frac{t}{n}) \\ &\leq n \exp[-\min\{c_i; c_i > 0\} \frac{t}{n}], \end{aligned}$$

pois $\frac{n}{a(t)} \leq \frac{n}{k} \leq n$.

Assim, (20) ocorre para t_ε tal que

$$n \exp[-\min\{c_i; c_i > 0\} \frac{t}{n}] < \varepsilon \quad \Leftrightarrow$$

$$t_\varepsilon \geq \frac{n}{\min\{c_i; c_i > 0\}} \log\left(\frac{n}{\varepsilon}\right). \quad \blacksquare$$

Ye [94] mostra que se o ponto inicial escolhido, no algoritmo de Karmarkar é o ponto $x^0 = \left(2^{-n}, \frac{1-2^{-n}}{n-1}, \dots, \frac{1-2^{-n}}{n-1}\right)^T$, então não se obtém em $O(\log(n/\varepsilon))$ iterações a iterada x^k satisfazendo

$$\frac{c^T x^k}{c^T x^0} \leq \varepsilon. \quad (3.22)$$

Mostraremos que o Algoritmo A aplicado ao problema (20), partindo do ponto x^0 acima, encontra uma ε -solução em uma iteração, com tamanho do passo

$$t_\varepsilon = \frac{n(n-1)}{(1-2^{-n}) \min\{c_i; c_i > 0\}} \log(n/\varepsilon).$$

Neste caso, em analogia à demonstração do Teorema 11 definimos

$$a(t) := 2^{-n} + \frac{1-2^{-n}}{n-1} \sum_{i=2}^n \exp\left[-\left(\frac{1-2^{-n}}{n-1}\right) c_i t\right],$$

então

$$\gamma_1(t) = \frac{2^{-n}}{a(t)}, \quad \gamma_i(t) = \frac{1-2^{-n}}{a(t)(n-1)} \exp\left[-\left(\frac{1-2^{-n}}{n-1}\right) c_i t\right]$$

e a função custo, ao longo da geodésica $\gamma(t)$,

$$h: \mathbf{R}_+ \rightarrow \mathbf{R}, \quad h(t) = -\frac{n-1}{1-2^{-n}} \cdot \frac{a'(t)}{a(t)} \quad (3.23)$$

Teorema 3.12. A função h , definida em (23), é monótona decrescente; além disso, o Algoritmo A, aplicado ao problema (20), encontra uma solução aproximada verificando

$$\frac{c^T x_k}{c^T x^0} < \varepsilon, \quad \text{para } \varepsilon > 0 \text{ dado,}$$

em uma iteração; uma estimativa do tamanho do passo pode ser dado por

$$t_\varepsilon = \frac{n(n-1)}{(1-2^{-n}) \min\{c_i; c_i > 0\}} \log(n/\varepsilon).$$

Prova: Análoga ao teorema anterior. ■

3.5. Convergência Global: Problemas Convexos em variedades de curvatura não-negativa

Nesta seção exploramos a convexidade para a obtenção da convergência global dos métodos geodésicos. Será no entanto essencial que a variedade possua curvatura não-negativa, de modo a se garantir uma propriedade geral dos métodos da Programação Não-Linear, conforme obtido, no \mathbf{R}^n , por Burachik [95] (Teorema 14 desta seção).

O lema abaixo será usado na prova de convergência dos algoritmos B e C .

Lema 3.2: Seja M um espaço métrico completo. Se a seqüência $\{x_k\}_{k \in \mathbf{N}}$, com $x_k \in M$, tiver ponto de acumulação \bar{x} satisfazendo

$$d^2(x_{k+1}, \bar{x}) \leq d^2(x_k, \bar{x}) + \delta_k \tag{3.24}$$

onde $\delta_k > 0$ e $\sum_{k=\ell}^{\infty} \delta_k$ é somável, então

$$\lim_{k \rightarrow \infty} x_k = \bar{x}.$$

Prova: Dado $\varepsilon > 0$, existe $n_1 \in \mathbf{N}$ tal que

$$\sum_{k=n_1}^{\infty} \delta_k < \frac{1}{2}\varepsilon^2 \quad \text{e} \quad d^2(x_{n_1}, \bar{x}) < \frac{1}{2}\varepsilon^2 \tag{3.25}$$

Para $k > n_1$, usando (24), temos

$$d^2(x_k, \bar{x}) \leq d^2(x_{n_1}, \bar{x}) + \sum_{j=n_1}^k \delta_j.$$

Agora, usando (25) na última desigualdade obtemos

$$d(x_k, \bar{x}) < \varepsilon \text{ para todo } k > n_1$$

o que mostra o lema. ■

O teorema seguinte, que generaliza o resultado similar de Kiwiel [83] para \mathbf{R}^n , é o resultado mais importante desta seção, pois com ele é possível provar convergência de algoritmos para funções convexas, não necessariamente diferenciáveis. Em Oliveira-Ferreira [95] desenvolvem algoritmos de sub-gradientes em Variedades Riemannianas para funções convexas não-diferenciáveis, usando o resultado abaixo para provar sua convergência.

Teorema 3.13: Seja $f: M \rightarrow \mathbf{R}$ convexa. Se a curvatura seccional de M é não-negativa, então para qualquer $y \in M$, vale:

$$d^2(x_{k+1}, y) \leq d^2(x_k, y) + \lambda_k^2 \|\text{grad } f(x_k)\|^2 + 2\lambda_k(f(y) - f(x_k)) \quad (3.26)$$

onde $\{x_k\}$ é a seqüência gerada pelo método gradiente e λ_k é o tamanho do passo.

Prova: Sejam $\gamma_i: [0, a_i] \rightarrow M$, $i = 1, 2$, geodésicas minimizantes normalizadas tais que $\gamma_1(0) = x_k$, $\gamma_1(a_1) = y$, $\gamma_2(0) = y$ e $\gamma_2(a_2) = x_{k+1}$; pelo Teorema 1.14 (item 1.14), temos

$$d^2(x_{k+1}, y) \leq d^2(x_k, y) + \lambda_k^2 \|\text{grad } f(x_k)\|^2 - 2d(x_k, y)\lambda_k \|\text{grad } f(x_k)\| \cos \theta \quad (3.27)$$

onde θ é o ângulo que os vetores $\dot{\gamma}_1(0)$, $-\text{grad } f(x_k)$ formam.

Como $\cos(\pi - \theta) = -\cos \theta$

$$\text{e} \quad \langle \text{grad } f(x_k), \dot{\gamma}_1(0) \rangle = \|\text{grad } f(x_k)\| \cos(\pi - \theta)$$

usando (27), segue

$$d^2(x_{k+1}, y) \leq d^2(x_k, y) + \lambda_k^2 \|\text{grad } f(x_k)\|^2 + 2\lambda_k d(x_k, y) \langle \text{grad } f(x_k), \dot{\gamma}_1(0) \rangle \quad (3.28)$$

do Teorema 2.6 (item 2.3.3) e (28) segue (25). ■

Vimos na seção 3.3 que o método gradiente, mesmo com a hipótese de gradiente ser Γ -lipschitziano - algoritmo C - somente obtivemos a convergência fraca da seqüência $\{x_k\}$ ao conjunto de pontos críticos de f com a hipótese de conjunto de nível limitado; nos próximos dois teoremas provaremos a convergência global das seqüências geradas pelos algoritmos B, C sem a hipótese de conjuntos de níveis limitados; por outro lado faremos hipótese sobre a curvatura de M e a convexidade de f .

Teorema 3.14: Seja $f: M \rightarrow \mathbf{R}$ convexa, de classe C^1 e gradiente Γ -Lipschitziano e M com curvatura não-negativa. Então a seqüência gerada pelo algoritmo C converge globalmente para um ponto de mínimo, caso exista.

Prova: Seja $a = \min_{x \in M} f(x) > -\infty$. Então $M^a \neq \emptyset$; seja $y \in M^a$. Pela desigualdade (26) e por se ter $y \in M^a$, chega-se a

$$d^2(x_{k+1}, y) \leq d^2(x_k, y) + \lambda_k^2 \|\text{grad } f(x_k)\|^2 \quad (3.29)$$

Usando que $\sum_{k=1}^{\infty} \lambda_k^2 \|\text{grad } f(x_k)\|^2 \leq \frac{1}{\beta}(f(x_0) - a)$ (veja demonstração do Teorema 3.9) e a desigualdade (29) temos que $d^2(x_k, y) \leq d^2(x_0, y) + \frac{1}{\beta}(f(x_0) - a)$; portanto, $\{x_k\}$ é limitada.

Pelo Teorema de Hopf-Rinow, temos que a seqüência $\{x_k\}$ tem ponto de acumulação x^* ; como a seqüência $\{f(x_k)\}$ é decrescente, veja (Teorema 9), pela continuidade da função f , segue-se que

$$\lim_{k \rightarrow \infty} f(x_k) = f(x^*) \quad \text{e} \quad f(x^*) \leq f(x_k). \quad (3.30)$$

Usando (30), (25) e (29), conclui-se que a seqüência $\{x_k\}$ tem ponto de acumulação x^* satisfazendo

$$d^2(x_{k+1}, x^*) \leq d^2(x_k, x^*) + \delta_k, \quad (3.31)$$

com $\delta_k := \lambda_k^2 \|\text{grad } f(x_k)\|^2$ e $\sum_{k=1}^{\infty} \delta_k$ somável.

Pelo Lema 2, temos

$$\lim_{k \rightarrow \infty} x_k = x^*. \quad (3.32)$$

Usando a desigualdade $0 < \delta_1 \leq \lambda_k$ e $\lim_{k \rightarrow \infty} \lambda_k \|\text{grad } f(x_k)\| = 0$, obtém-se $\lim_{k \rightarrow \infty} \|\text{grad } f(x_k)\| = 0$.

De (32) e da continuidade do gradiente, temos

$$\text{grad } f(x^*) = 0.$$

Finalmente da convexidade de f , deduz-se que x^* é ponto de mínimo. ■

Vimos, pela observação após o Teorema 8, que qualquer ponto de acumulação da seqüência gerada pelo Algoritmo B é ponto crítico da função f ; mostraremos que para f convexa e M com curvatura positiva, a seqüência converge globalmente para um ponto de mínimo, caso exista.

Teorema 3.15: Seja $f: M \rightarrow \mathbf{R}$ convexa, de classe C^1 e M com curvatura não-negativa. Então a seqüência gerada pelo Algoritmo B converge globalmente para um ponto de mínimo de f , caso exista.

Prova: Pela observação acima, basta provar que a seqüência é convergente, caso f admita mínimo.

Seja $a := \min_{x \in M} f(x)$, $y \in M^a$. Pelo Teorema 13, $y \in M^a$ temos:

$$d^2(x_{k+1}, y) \leq d^2(x_k, y) + \lambda_k^2 \|\text{grad } f(x_k)\|^2. \quad (3.33)$$

Do modo como λ_k é escolhido, temos que

$$\sum_{k=0}^m \lambda_k^2 \|\text{grad } f(x_k)\|^2 \leq \bar{t} \sum_{k=0}^m \lambda_k \|\text{grad } f(x_k)\|^2 \leq \frac{\bar{t}}{\alpha} (f(x_0) - a);$$

da desigualdade (33), segue que

$$d^2(x_k, y) \leq d^2(x_0, y) + \frac{\bar{t}}{\alpha} (f(x_0) - a),$$

isto é,

$\{x_k\}$ é limitada.

Pelo Teorema de Hopf-Rinow e convexidade de f segue que a seqüência $\{x_k\}$ tem ponto de acumulação $x^* \in M^a$. Novamente da desigualdade (33) e do Lema 2, chega-se

$$\lim_{k \rightarrow \infty} x_k = x^*. \quad \blacksquare$$

CONCLUSÃO

Desenvolvemos uma teoria de convergência em que os resultados se assemelham aos clássicos da otimização irrestrita, o que corresponde a nosso objetivo inicial.

Uma extensão desta tese é a resolução dos problemas abaixo, fazendo uso dos resultados de convergência apresentados nos Capítulos 1 e 3.

Problema 1. Minimização de uma função quadrática convexa no simplex, isto é,

$$\min \frac{1}{2} x^T Q x + b^T x + c$$

s.a.

$$\sum_{i=1}^n x_i = 1 \tag{1}$$

$$x_i \geq 0, \quad i = 1, \dots, n$$

onde Q é uma matriz simétrica semi-definida positiva, $b \in \mathbf{R}^n$ e $c \in \mathbf{R}$.

O problema (1) aparece no desenvolvimento de métodos de descida para problemas de otimização não-diferenciável, Kiwiel [86], e é de grande importância para a indústria de mineração, Szidarovszky [87].

Além da teoria específica, espera-se que a implementação do algoritmo B leve a resultados atraentes.

Problema 2. Dada a matriz Q simétrica, encontrar $0 \neq x \in \mathbf{R}^n$, $\lambda \in \mathbf{R}$ que resolva

$$Qx = \lambda x \tag{2}$$

Uma solução para (2) pode ser obtida como solução da minimização de uma função

quadrática convexa restrita à esfera $S^{n-1} = \{x \in \mathbf{R}^n; \sum_{i=1}^n x_i^2 = 1\}$:

$$\min \frac{1}{2} x^T Q x$$

s.a.

$$\sum_{i=1}^n x_i^2 = 1$$

A esfera S^{n-1} é o modelo de variedade Riemanniana completa (compacta) de curvatura seccional constante positiva ($K \equiv 1$), do Carmo; além disso, dados $p \in S^{n-1}$ e $v \in T_p S^{n-1} \|v\| = 1$, a geodésica com essas condições iniciais é dada por

$$\gamma(t) = p \cos t + v \operatorname{sen} t, \quad t \in \mathbf{R},$$

$$(\gamma''(t) = -\gamma(t)).$$

Novamente, podemos aplicar o algoritmo B ou então calcular a constante de Lipschitz do campo gradiente e aplicar o algoritmo C.

Problema 3. O problema

$$\min f(x)$$

s.a

$$x \geq 0 \tag{3}$$

corresponde a um modelo de interesse, seja em termos de aplicações, Iusem [95], seja em termos de desenvolvimento de métodos, Poliak [87].

No ortante positivo, temos as geodésicas explícitas, curvatura nula e expressão fechada para o hessiano, o que permite desenvolver métodos de Newton (2ª ordem) ou então aplicar os algoritmos do Capítulo 3.

Finalmente, pensamos também em estabelecer uma teoria local, isto é, obter uma “lei dos cossenos” onde se tenha informação de curvatura não-negativa em uma vizinhança do ponto ótimo e, com isso, estender os resultados de convergência.

BIBLIOGRAFIA

- Adler, I., Monteiro, R.D.C.** Limiting Behavior of the Affine Scaling Continuous Trajectories for Linear Programming Problems. *Contemporary Mathematics* 144 (1990) 189–211.
- Anstreicher, K.M.** On the Performance of Karmarkar Algorithm over a Sequence of Iterations. *SIAM J. Optimization* 1, 1 (1991) 22–29.
- Bayer, D.A., Lagarias, J.C.** The Non-Linear Geometry of Linear Programming, I. Affine and Projective Scaling Trajectories. II. Legendre Transform Coordinates and Central Trajectories. *Trans. Amer. Math. Soc.* 314 (1989) 499–581.
- BISHOP, R.L., O’Neill, B.** Manifolds of Negative Curvature. *Transactions of the American Mathematical Society*, 145 (1969) 1–49.
- Burachik, R., Grana Drummond, L.M., Iusem, A.N., Svaiter, B.F.** Convergence of the Steepest Descent with Inexact Line Searches. *Optimization* 32 (1995) 137–146.
- Dikin, I.I.** Iterative Solution of Problems of Linear and Quadratic Programming. *Soviet Math. Dokl.* 8 (1967) 674–675.
- Diener, I.** On the Global Convergence of Path-Following Methods to Determine all Solutions to a System of Nonlinear Equations. *Mathematical Programming* 39 (1987) 181–188.
- Do Carmo, M.P.** *Differential Geometry of Curves and Surfaces.* Prentice-Hall, New Jersey (1976).
- Do Carmo, M.P.** *Geometria Riemanniana.* 2ª edição, IMPA-CNPq, Rio de Janeiro (1988).
- Eggermont, P.P.B.** Multiplicative Iterative Algorithms for Convex Programming. *Linear Algebra and Its Applications* 130 (1990) 25–42.
- Gabay, D.** Minimizing a Differentiable Function over a Differential Manifold. *JOTA* 37, 2 (1982) 177–218.

- Gonzaga, C.C.** Path-Following Methods for Linear Programming. *SIAM Review* 34, 2 (1992) 167–224.
- Iusem, A.N.** An Interior Point Multiplicative Method for Optimization Under Positivity Constraints. *Acta Applicandae Mathematicae* 38 (1995) 163–184.
- Karmarkar, N.** A New Polynomial - Time Algorithm for Linear Programming. *Combinatorica* 4 (1984) 373–395.
- Karmarkar, N.** Riemannian Geometry Underlying Interior - Point Methods for Linear Programming. *Contemporary Mathematics* 144 (1990) 51–75.
- Kiwiel, K.C.** An Aggregate Subgradient Method for Nonsmooth Convex Minimization. *Mathematical Programming* 27 (1983) 320–341.
- Kiwiel, K.C.** A Method for Solving Certain Quadratic Programming Problems Arising in Nonsmooth Optimization. *IMA Journal of Numerical Analysis* 6 (1986) 137–152.
- Lagarias, J.C.** The Nonlinear Geometry of Linear Programming. III. Projective Legendre Transform Coordinates and Hilbert Geometry. *Transaction of The American Mathematical Society* 320, 1 (1990) 193–226.
- Luenberger, D.G.** The Gradient Projection Method Along Geodesic. *Management Science* 18, 11 (1972) 620–631.
- Luenberger, D.G.** Introduction to Linear and Nonlinear Programming. Addison-Wesley Publishing Company, Massachusetts [1984].
- Meggido, N., Shub, M.** Boundary Behavior of Interior Point Algorithms in Linear Programming. *Mathematics of Operations Research* 14 (1989) 97–146.
- Mercuri, F., Rigas, A.** Curvatura e Topologia. Escola de Geometria Diferencial, UFC, Fortaleza (1978).
- Milnor, J.** Morse Theory. Princeton University Press, New Jersey (1968).

- Milnor, J.** Morse Theory. Princeton University Press, New Jersey (1968).
- Oliveira, P.R., Xavier, J.C.N.** Elementos de Geometria Riemanniana. Publicações Técnicas, COPPE, UFRJ (1994).
- Oliveira, P.R., Ferreira, O.P.** Algoritmo Subgradiente em Variedade Riemanniana. Publicações Técnicas, COPPE, UFRJ (1995).
- Poliak, B.** Introduction to Optimization. Optimization Software, New York (1987).
- Rapcsák, T.** Geodesic Convexity in Nonlinear Optimization. JOTA 69, 1 (1991).
- Szidarovszky, F., Baafi, E.Y., Kim, Y.C.** Kriging Without Negative Weights. Mathematical Geology 19, 6 (1987).
- Ye. Y., Ji, J.** A Complexity Analysis for Interior - Point Algorithms Based on Karmarkar's Potential Function. SIAM J. Optimization 4, 3 (1994) 512–520.
- Wu, S.** Convergence Properties of Descent Methods for Unconstrained Minimization. Optimization 26 (1992) 229–237.